PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, G01N 33/50 — A1

(11) International Publication Number: **WO 98/45476**

(43) International Publication Date: 15 October 1998 (15.10.98)

(21) International Application Number: PCT/GB98/00869

(22) International Filing Date: 8 April 1998 (08.04.98)

(30) Priority Data:
9707141.9 8 April 1997 (08.04.97) GB

(71) Applicant *(for all designated States except US)*: INSTITUTE OF FOOD RESEARCH [GB/GB]; Norwich Laboratory, Norwich Research Park, Colney, Norwich NR4 7UA (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: SCHWEIZER, Michael [DE/GB]; 34A Lynch–Green, Hethersett, Norwich NR9 3JT (GB).

(74) Agent: THOMAS, Philip; Eric Potter Clarkson, Park View House, 58 The Ropewalk, Nottingham NG1 5DD (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: BIOLOGICAL ASSAY FOR TESTING THE CARCINOGENIC PROPERTIES OF A SUBSTANCE

(57) Abstract

An assay for testing the carcinogenic properties of a test substance comprising: (i) introducing into cells a reporter gene expression vector comprising a repetitive DNA sequence which exhibits instability in cancer cells, whereby instability of the repetitive DNA sequence affects expression of the reporter gene; (ii) exposing the resulting cells to the test substance; and (iii) determining whether the test substance is carcinogenic or anticarcinogenic by comparing the frequency of reporter gene expression in the resulting cells with the frequency of reporter gene expression in cells which have not been exposed to the test substance. The invention also provides expression vectors suitable for use in the assay methods.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

BIOLOGICAL ASSAY FOR TESTING THE CARCINOGENIC PROPERTIES OF A SUBSTANCE

The present invention relates to a biological assay and assay reagents for testing the carcinogenic properties of a test substance. The assay is useful for screening potential anti-cancer drugs as well as for testing the carcinogenic properties of food components.

DNA repair is an essential process in all organisms from prokaryotes to eukaryotes. Defective DNA repair in higher eukaryotes such as humans is an important factor in the aetiology of both hereditary and sporadic carcinomas. According to the current model of carcinogenesis, initiation is a single-cell event which leads to the development of a precancerous lesion by clonal expansion. Progression to an invasive tumour is a prolonged process requiring the acquisition of several further mutations in genes controlling cell proliferation and differentiation. As the spontaneous mutation rate in normal cells is relatively low it suggests that an early event in the pathway of tumourgenesis is a mutation that confers a so-called "mutator" phenotype. This postulates that defects in DNA repair occur at an early stage in the sequence of events and favour accelerated tumour progression. This hypothesis is corroborated by the recent link between cancer, microsatellite instability and mutations in the genes encoding the mismatch repair machinery. Mismatch repair involves the processing of incorrectly paired nucleotides which can occur as a result of normal DNA metabolism. It also plays an important role in the recognition and correction of unpaired loop structures which form during the replication of highly repetitive regions of DNA known as microsatellite DNA. If left unrepaired these loop-structures can lead to frame-slippage and potential loss of gene function.

Hereditary non-polyposis cancer (HNPCC) is a cancer of the colon characterised by microsatellite instability. HNPCC is an autosomal dominant disease in which multiple members of a family suffer early onset colon cancer in the absence of polyp formation. Of the HNPCC tumours that exhibit microsatellite instability over 50% contain mutations in the HNPCC-linked *hMSH2* gene on chromosome 2 and about 20-30% contain mutations in the HNPCC-linked *hMLH1* gen on chromosome 3 (Umar, A. & Kunkel, T.A., 1996).

HNPCC-Lynch syndromes I and II is a common cancer predisposition syndrome that is autosomal dominant in nature. Lynch I families suffer early onset colorectal cancer, while Lynch II kindreds are also susceptible to extra colonic epithelial tumours of the endometrium, ovary, stomach, small intestine, kidney and ureter. The link between microsatellite instability and cancer is demonstrated by the fact that a subset of sporadic colon cancers and the majority of tumours occurring in HNPCC patients contain frequent mutations in the simple microsatellite sequences (A)n, (GGC)n, or (CA)n. These mutations seem to be tumour specific with each cell containing thousands of microsatellite mutations. Studies involving microsatellite instability show that it is present in a significant number of sporadic tumours including colorectal (12-28%), endometrial (17-23%), stomach (18-39%), ovarian (16%), cervical (15%), pancreatic (67%), oesophageal adenoma (22%), squamous cell skin (50%), and small-cell lung cancer (45%) (Eshleman, J.R. & Markowitz, S.D. 1995; Aaltonen *et al.* 1993; Merlo *et al.* 1994; Mironov *et al.* 1994; Orth *et al.* 1994; Modrich, M. & Lahue, R. 1996).

Although not fully understood, the mismatch repair pathways of both lower and higher eukaryotes share extensive homology. For example, homologues of the human mismatch repair pathway exist in *Saccharomyces cerevisiae*.

Henderson and Petes, *Mol. and Cell. Biol.*, June 1992, **12**, No. 6, p.2749-2757 have constructed reporter gene expression vectors for studying spontaneous frameshift mutations. The vectors are based on a plasmid having the LEU2 promoter and the first 12 codons of the yeast LEU2 protein fused to the eighth codon of the *E. coli* β-galactosidase gene (lacZ). They inserted various oligonucleotides containing simple repetitive DNAs into the *Bam*HI site near the beginning of the β-galactosidase gene.

The inserts did not shift the reading frame so that β-galactosidase expression occurred unless a frameshift mutation occurred in the host (yeast) cells. Such events were visible as white colonies when the cells were grown on a medium containing Xgal.

Similar reporter gene expression vectors were constructed in which the repetitive DNA tract was inserted upstream of the URA3 gene. On a medium containing 5-fluoro-orotic acid (5-FOA) frameshift mutations were detected as URA3-(5-FOA^R) cell colonies.

Levinson and Gutman (1987), *Nuc. Acids Res.*, **15**, p.5323-5339 also used a reporter gene expression vector to study frameshift mutations in the prokaryote *E. coli* K12. The vector was based on bacteriophage M13 and contained short poly-CA/TG tandem repeats linked to the lacZ gene which encodes β-galactosidase.

Strand *et al*, *Nature*, **365**, p.274-276, 16 September 1993, used a reporter gene expression vector comprising a yeast promoter fused to a β-galactosidase gene that contained a 29-base pair out-of-frame poly (GT) tract in the coding sequence to study repetitive DNA tract instability. They found that mutations in any three yeast genes involved in DNA mismatch repair (PMS1, MLH1 and MSH2) lead to 100- to 700-fold increases in repetitive tract instability, whereas mutations that eliminate the proof reading function of DNA polymerases have little effect.

According to a first aspect of the invention there is provided an assay for testing the carcinogenic properties of a test substance comprising: (i) introducing into cells a reporter gene expression vector comprising a repetitive DNA sequence which exhibits instability in cancer cells, whereby instability of the repetitive DNA sequence affects expression of the reporter gene; (ii) exposing the resulting cells to the test substance; and (iii) determining whether the test substance is carcinogenic or anti-carcinogenic by comparing the frequency of reporter gene expression in the resulting cells with the frequency of reporter gene expression in cells which have not been exposed to the test substance.

The term "carcinogenic properties" is intended to embrace the ability of the test substance to inhibit cancer as well as to cause cancer, that is, the term embraces both carcinogenic and anti-carcinogenic properties.

By "instability" we mean a change in the size of the DNA sequence, normally by additions or deletions that are not a multiple of 3 bp. Such changes in size alter the reading frame for transcription of adjacent genes

and are known as frameshift mutations.

Preferably, the repetitive DNA sequence (often referred to as microsatellite DNA) comprises a poly d(AC/TG) tract and/or a poly d(GT/CA) tract, although the tract may comprise a single nucleotide eg. poly d(G) or poly d(A).

It will be appreciated that the length of the repetitive DNA tract can be varied and is preferably selected according to the length of the repetitive sequence identified as being unstable in the cancer cell of interest. However, the length of the repetitive sequence is conveniently 8 to 60 nucleotides, more preferably 16 to 32 and especially 16.

The term "reporter gene expression vector" is intended to cover any vector into which a reporter gene has been inserted so that, on introduction into a suitable host cell, the reporter gene will be transcribed and translated to produce the protein product of the reporter gene.

A skilled person will appreciate that the expression vector can be provided in a variety of forms eg. a plasmid, a 'phage or a virus.

Preferably the reporter gene expression vector comprises a promoter region of a gene which is normally expressed in the host cell fused to a sequence encoding a reporter gene product which can be expressed in the host cell. Preferably a repetitive DNA sequence which exhibits instability in cancer cells is inserted downstream of the promoter region into the open-reading frame of the reporter gene sequence. The open-reading frame (ORF) will be understood by skilled persons to mean a DNA

sequence which contains a series of triplets coding for amino acids without any termination codons.

The insertion preferably "knocks" the promoter/reporter gene fusion out-of-frame (+1 or -1 reading frame) so that the reporter gene is not expressed. Hence expression of the reporter gene only occurs if the inserted repetitive sequence changes size (exhibits instability) so that the correct reading frame of the reporter gene is established.

Alternatively, the insertion does not knock the promoter/reporter gene fusion out of frame so that expression of the reporter gene occurs unless the inserted repetitive sequence exhibits instability so that it knocks the promoter/reporter gene out-of-frame.

It is within the knowledge of skilled persons to select a variety of promoter sequences and reporter gene sequences which can be used in a given host cell be it yeast, human or bacterial. The reporter gene expression vectors disclosed by Henderson and Petes (1992), Levinson and Gutman (1987), and Strand *et al* (1993) are incorporated herein by reference.

Preferably the reporter gene expression vector is provided in the form of a low or high copy number plasmid, or an integrative plasmid, that is, a plasmid which lacks a host cell origin of replication and must therefore be integrated into the host cell genome for stable maintenance in the host cells.

Preferably the reporter gene comprises a gene whose expression product

gives rise to a visible change in the host cell. For example the gene product may produce a colour change or fluorescence. A particularly preferred reporter gene system comprises the lacZ gene which encodes the enzyme β-galactosidase. β-galactosidase expression can be detected as a blue colour in colonies growth on a medium containing Xgal. Colonies which do not express β-galactosidase appear as white colonies. Hence, use of the vector comprising the lacZ reporter gene according to a preferred embodiment provides a simple blue/white colour test for screening the carcinogenic effect of a test substance.

Preferably the cells used in the assay are eukaryotic cells, preferably yeast cells or human cells, and especially eukaryotic cells which have a defect in repetitive DNA instability repair mechanisms, especially the mismatch repair pathway. As homologues of the human mismatch repair pathway exist in yeast such as *Saccharomyces cerevisiae* this single cell eukaryote provides an ideal model for studying the effects of test substances eg. dietary constituents on DNA repeat instability in humans.

Of course human cell lines can be used directly in the assays of the invention, the cell lines being derived from the cancer of interest eg. human colorectal cancer, especially hereditary non-polyposis cancer (HNPCC). Although not preferred, the assay may use prokaryotic cells, conveniently bacterial cells such as *Escherichia coli*.

In a second aspect the invention provides an assay comprising testing the carcinogenic properties of a test substance using yeast cells according to the first aspect of the invention; and further testing the test substance using human cells according to the first aspect of the invention.

As yeast cells are easier to work with than human cells the first step of the assay according to the second aspect of the invention provides a primary screening step which can be used to screen out test substances exhibiting undesirable carcinogenic properties before the second step whose results have greater significance in humans.

In a third aspect the invention provides an assay for testing the carcinogenic properties of a test substance comprising the steps of: (i) screening a test substance using a reporter gene expression vector according to the first aspect of the invention; and (ii) repeating the assay using cells containing a high, medium, low or single copy number vector, the high, medium, low or single copy number vector being selected depending on the frequency of repetitive DNA instability measured in step (i).

Preferably the reporter gene expression vector in step (i) is contained in the cells at a high copy number.

The assay according to the third aspect of the invention is particularly useful when the frequency of instability in step (i) is very high in cells which have not been exposed to the test substance. By selecting medium, low or single copy number vectors the background instability can be reduced as desired to make the assay more sensitive.

Preferred non-limiting embodiments of the invention will now be described with reference to the accompanying drawings in which:

Figure 1 shows the steps of a preferred assay according to the first aspect of the invention. The assay utilises a preferred assay plasmid of the

invention;

Figure 1A shows the high copy assay plasmid pKa3-9(n) as previously described. The unique *EcoRI* site into which the respective tracts were cloned is underlined. Also underlined is the unique *Stu*I site situated within the *URA3* gene, used for the insertion of the KanMX4 cassette (cf. text for further details).

Figure 1B shows the high copy assay plasmid pKa3-9(n)KanMX4 utilising the dominant selectable marker KanMX4. Both the *EcoRI* site into which the respective tracts were cloned and the additional *EcoRI* site derived from the KanMX4 cloning step are underlined (cf. text for further details).

Figure 2 shows a yeast/*E. coli* shuttle vector suitable for fusing yeast promoter and coding sequences to the lacZ gene of *E. coli*;

Figure 2A shows the centromere-based assay plasmid pKaCEN(n) as previously described. The unique *EcoRI* site into which the repetitive tracts were cloned is underlined. Also underlined is the unique *Stu*I site situated within the *URA3* gene, used for the insertion of the KanMX4 cassette (cf. text for further details).

Figure 2B shows the centromere-based assay plasmid pKaCEN(n)KanMX4 utilising the dominant selectable marker KanMX4. Both the *EcoRI* site into which the repetitive tracts were cloned and the additional *EcoRI* site derived from the KanMX4 cloning step are underlined (cf. text for further details).

Figure 3 shows a low copy number vector containing sequences from yeast centromere VI;

Figure 4 shows a vector which lacks a yeast origin of replication, so that it must be integrated into the yeast genome for stable maintenance;

Figure 5 shows a preferred high copy number reporter gene

expression vector according to the invention;

Figure 6 shows a preferred low copy number reporter gene expression vector according to the invention;

Figure 7 shows a preferred single copy integrative reporter gene expression vector according to the invention;

Figure 8 is a schematic representation which shows integration of the vector of Figure 7 at the URA3 locus of *S. cerevisiae* strain YN94-1;

Figures 9 and 10 show the results of experiments to confirm insertion of the integrative vector of Figure 7 into the *S. cerevisiae* strain as illustrated in Figure 8;

Figure 11 shows the synthetic pathway for the polyamines putrescine, spermidine and spermine in eukaryotes;

Figure 12 is a schematic representation of frame-slippage in the preferred vectors leading to β-galactosidase reporter gene expression;

Figures 13a and 13b shows a Southern blot analysis of MSH2 genomic DNA (Figure 13a) and the disruption scheme (Figure 13b);

Figure 14 shows a Southern blot analysis of wild type and disrupted MLH1 gene;

Figure 15 shows a vector incorporating the SV40 promoter upstream of the luciferase gene;

Figures 15a and 15b show Southern blot analysis of YN97-150 (*msh :: KanMX4*);

Figure 16 shows an EBV-based vector for stable expression of DNA in human host cells;

Figures 16a and 16b show Southern blot analysis of YN97-167 (*rth1 :: KanMX4*);

Figure 17 shows the effect of a known carcinogen on the instability of the high copy number expression vector of Example 1(I).

Figures 17a and 17b show Southern blot analysis of YN98-3 (pol 130-104 :: *LEU2*);

Figure 18 is a schematic outlining biosynthesis of purine nucleotides, from Jones and Fink, 1982.

MATERIALS

1. PREFERRED YEAST STRAINS

YN94-1: *MATa, ade2-1, his3-11, leu2-3, 112, trp1-1, ura3-1, can1-100.*

2. PREFERRED *E. COLI* STRAINS

DH5α: F,_80*lac*Z△M15 △(*lac*ZYA-argF)U169 *deo*R *rec*A1 *end*A1 *hds*R17(r_K-, m_K+) *pho*A *sup*E44 γ- *thi*-1 *gyr*A96 *rel*A1.

JM109: e14-(McrA-) *rec*A1 *end*A1 *gyr*A96 *thi-1 hds*R17 (r_K-m_K+) *sup*E44 *rel*A1 △(*lac-pro*AB) [F' *tra*D36 *pro*AB *lacl*qZ△M15].

RR1: F-, *hsd*S20 (r_B-, m_B-), *sup*E44, *ara*14, *pro*A2, *rps*L20 (str), *syl*-5, *mlt*-5, *sup*E44, γ-.

3. PREFERRED MEDIA

SCD minimal/defined medium

20g	glucose
1.7g	yeast nitrogen based - without amino acids and ammonium sulphate (Difco, Detroit, USA)
5g	ammonium sulphate
25ml	"drop-out" mix
20g	agar (if plates are required)

Made up to 1 litre with glass dist. H_20 and autoclaved at 121°C for 15 minutes.

"Drop-out" mix (40x):

0.2g	uracil	0.4g	lysine	0.1g	adenine
0.2g	arginine	0.1g	methionine	0.4g	tyrosine
0.1g	histidine	0.6g	phenylalanine	0.4g	tryptophan
0.6g	isoleucine	0.5g	threonine	0.6g	leucine

"Drop-out mix" lacking uracil (SCD-U) was used for selection of assay plasmids. Food components at various concentrations were added aseptically before the plates were poured.

LB (Luria-Bertani) Medium (for growth of *E. coli*)

10g Bacto-tryptone (Difco)
5g Bacto-yeast extract (Difco)
10g NaCl
15g Agar (if plates are required)

Make up to 1 litre with glass dist. H_2O, Adjust pH to 7.5 with NaOH and autoclave at 121°C for 15 minutes.

For Amp^R selection, ampicillin is added to a final conc. of 100µg/ml when the media has cooled to approximately 55°C.

4. PREFERRED REPETITIVE OLIGONUCLEOTIDES USED

IN THE CONSTRUCTION OF ASSAY PLASMIDS

(purchased from Oswel DNA service, Southampton)

Oligonucleotide 1

(47 nt, poly $d(TG)_{16}$, with an *Eco*RI restriction site overhang at its 5' end and a *Smal* site, indicated by underlining.)

Concentration. 1413mg/ml

T_m 0.1 M Na+ 74.1°C

Smal

↓

5'-AATTGCCCGGGCTGTGTGTGTGTGTGTGTGTGTGTG TGTGTGCCG-3'

Oligonucleotide 2

(47 nt, poly $d(AC)_{16}$, with an *Eco*RI restriction site overhang (underlined) at its 3' end).

Concentration 778mg/ml

T_m 0.1 M Na+ 74.1°C

5'-CGGGCCCGACACACACACACACACACACACACACACAC GGCTTAA-3'

5. **PREFERRED PRIMER SEQUENCES**

LacZ Reverse: $_{529}$ 5'-AAGGGGGATGTGCTGCAAGG-3' $_{509}$

YEp Forward: $_{8271}$ 5'-GCAGCGAGTCAGTGAGCGAGG-3 $_{8291}$

The numbers in subscript indicate the positions of annealing on plasmid YEp356R (cf. Fig. 2).

PRS3 Forward: $_{-50}$ 5'-GATTCATACTCTTTTTTCTACG -3' $_{-81}$

The numbers in subscript indicate the positions of annealing relative to the ATG start codon of *PRS3*.

Repeat tracts of varying sequence and length were synthesized and inserted into the unique *Eco*RI site of pSS3-9 as previously described.

Poly $(AC)_{15}A$ assay tract

(shifts *PRS3'/LacZ* fusion into +1 reading frame with respect to start ATG of *PRS3*)

Oligonucleotide 1

(poly $d(AC)_{15}A$, with an *Eco*RI restiction site overhang at its 5'-end and a *Sma*I site at its 3'-end both indicated by underlining, giving a total length of 46 nts).

5'-AATTCGGACACACACACACACACACACACACACACAGCCC GGGC-3'

Oligonucleotide 2

(poly $d(TG)_{15}T$, with an *Eco*RI restiction site overhang at its 5'-end and a *Sma*I site at its 3'-end both indicated by underlining, giving a total length of 46 nts).

5'-AATTGCCCGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT CCG-3'

Poly (C_{1-3} A) telomeric assay tract

In *Saccharomyces cerevisiae* the sequence poly ($C_{1-3}A$) is found at or near the telomeres (Shampay *et al.* 1984). To examine the stability of these sequences a 65 bp telomeric tract was inserted into the *Eco*RI site of plasmid pSS3-9 as described.

(insertion shifts *PRS3'/LacZ* fusion into -1 reading frame with respect to start ATG)

Telomeric oligonucleotide 1

(poly $d(C_{1-3}A)$, with an *Eco*RI restriction site overhang at its 5'-end and a *Sma*I site at its 3'-end both indicated by underlining, giving a total length of 80 nts).

5'-AATTCGGACCACACCACCACACCCACACCACACCACACACC CACACACACACCACCACCCACACACCACACAGCCCGGGC-3'

Telomeric oligonucleotide 2

(80 nt, poly d(G_{1-3}T), with an *Eco*RI restriction site overhang at its its 5'-end and a *Sma*I site at its 3'-end both indicated by underlining, giving a total length of 80 nts).

5'-AATTGCCCGGGCTGTGTGGTGTGTGGGTGGTGGTGTGTGTG
TGGGTGTGTGGTGTGGTGTGGGTGTGGTGGTGTGGTCCG-3'

Poly $(A)_{20}$ tract

(shifts *PRS3'/LacZ* fusion into -1 reading frame with respect to start ATG)

Oligonucleotide 1

(poly d$(A)_{20}$, with an *Eco*RI restriction site overhang at its 5'-end and a *Sma*I site at its 3'-end both indicated by underlining, giving a total length of 35 nts).

5'-AATTCGGAAAAAAAAAAAAAAAAAAAAGCCCGGGC-3'

Oligonucleotide 2

(35 nt, poly d$(T)_{20}$, with an *Eco*RI restriction site overhang at its its 5'-end and a *Sma*I site at its 3'-end both indicated by underlining, giving a total length of 35 nts).

5'-AATTGCCCGGGCTTTTTTTTTTTTTTTTTTTTCCG-3'

During the cloning of repetitive tracts into pSS3-9 a number of aberrant events occurred resulting in assay plasmids containing repetitive tracts as follows. Each of these tracts were subcloned into the centromeric and integrative vectors as previously described and incorporated into the assay system.

tract	Reading frame
$(AC)_{16}$	- 1
$(AC)_{41}(A)_2(C)_2(AC)_4A$	- 1
$(AC)_{11}$	+1
$(AC)_{12}A$	+1
$(AC)_{15}A$	+1

Construction of assay plasmids utilising the dominant selectable marker *KanMX4*.

Loss of growth selection may be encountered when yeast strains containing the plasmid-borne *URA3* marker are grown on a non-selective medium. This can occur when whole food extract containing growth-sustainable amounts of uracil are added to the medium. This problem can be overcome using a plasmid-borne KanMX4 dominant selectable marker consisting of the *E. coli* transposon T*n*903 fused to the transcriptional control sequences of the *TEF* gene of the filamentous fungus *Ashbya gossypii*. This hybrid molecule permits the efficient selection of yeast transformants resistant to geneticin (G418) (Wach *et al.* 1994).

Assay plasmids as described above containing the *URA3* gene (cf Figs. 1A and 2A) were linearised by digestion at the unique *Stu*I restriction enzyme

site (Promega), (+436 relative to the URA3 start codon). 40 ng of this linearised vector was ligated to 200 ng of a gel-purified 1481 bp *Eco*RV/*Sma*I fragment from pFA6-KanMX4 (Wach et al. 1994) containing the KanMX4 cassette. An aliquot of this ligation mixture was then transformed into the *E. coli* strain RR1 and colonies growing on kanamycin were selected. Plasmid DNA was isolated from several kan-resistant colonies and subjected to restriction analysis. (The insertion of the KanMX4 cassette introduces a second *Eco*RI site into the assay plasmids as illustrated by comparing Figure 1A with 1B and 2A with 2B.)

When assayed for repeat instability in *S. cerevisiae* these G418 selectable plasmids gave similar frequencies of frame-slippage to that measured for the original constructs relying on uracil selection. (cf. Table 1a).

CONSTRUCTION OF MISMATCH REPAIR MUTANTS

(a) *MSH2*

The *MSH2* gene of *Saccharomyces cerevisiae* (cf. Fig. 1) is one of several genes that share extensive homology with the bacterial *MutS* gene. Located on chromosome XV it encodes a protein of 109kda. Like the MutS protein, the MSH2p binds selectively to DNA containing mispairs and substrates containing up to 14 extra bases. Strains that contain mutated *MSH2* genes have strongly elevated rates of spontaneous mutations and exhibit microsatellite instability (cf. Table 1).

Disruption of *MSH2*

The *MSH2* gene of *Saccharomyces cerevisiae* TN94-1 was disrupted with *LEU2* (Fig. 13). Plasmid pRhB113 (Rhona Borts, Yeast Genetics, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU) containing the *MSH2* gene disrupted with *LEU2* at the *SnaBI* site was digested to completion with restriction enzyme *Spe*I in buffer **REact**1 (cf. Materials and Methods). The digestion mixture was transformed directly into *S. cerevisiae* YN904-1 (cf. Materials and Methods). Resulting transformants were then screened by Southern hybridization for the presence of the disrupted *MSH2* gene (Figures 13a and b).

Figure 13a. Southern blot analysis of YN97-10 (*msh2::LEU2*)

Southern blot analysis of *MSH2* genomic DNA (Fig. 13a) and corresponding disruption scheme (cf. Fig. 13b). Lane (1) *Hin*dIII digested control YN94-1 DNA. Two bands are seen (1650 and 2510 bp corresponding to the wild type *MSH2* gene on chromosome 15). Lane (2) *Hin*dIII digestion of genomic DNA containing disrupted *MSH2*. The gel was probed with a 1662 bp *Eco*RI fragment derived from pRhB113. A shift in the 2510 bp band is seen to 4710 bp corresponding to the insertion of *LEU2*. The disrupted *msh2::LEU2* strain was then assayed for the mutator phenotype (cf. Table 1).

(b) *MLH1*

The *MLH1* gene of *Saccharomyces cerevisiae* is one of several genes that shares extensive homology with the bacterial *MutL* gene. It is believed that MLH1p forms an interaction with MSH2p during the initiation of DNA mismatch repair in yeast. Yeast strains that contain mutated *MLH1*

genes have strongly elevated rates of spontaneous mutations and also exhibit microsatellite instability (cf. Table 1). The *MLH1* gene of *Saccharomyces cerevisiae* YN94-1 was disrupted with *LEU2*. Plasmid pREd182 (Rhona Borts, Yeast Genetics, Institute of Molecular Medine, John Radcliffe Hospital, Oxford OX3 9DU) containing the *MLH1* gene disrupted with *LEU2* was digested to completion with the restriction enzymes *Sac*I/*Bam*HI in buffer REact 3 (cf. Materials and Methods). This digestion mixture was transformed directly into *S. cerevisiae* YN94-1 (cf. Materials and Methods). Resulting transformants were then screened by Southern hybridization for the presence of the disrupted *MLH1* gene on chromosome XIII (Fig. 14).

Construction of mismatch repair pathway mutants for use in the assay system.

(c) *MSH3*

The *MSH3* gene of *Saccharomyces cerevisiae* is another gene that shares extensive homolgy with the bacterial *MutL* gene. MSH3p forms a heterodimer with MSH2p during the initiation of insertion/deletion mismatch repair. Yeast strains mutant for *MSH3* exhibit a less profound increase in microsatellite instability as compared to *MSH2* and *MLH1* mutant strains and have slightly elevated levels of spontaneous mutations (Strand *et al.* 1995).

Disruption of *MSH3*

The *MSH3* gene of *Saccharomyces cerevisiae* YN94-1 was disrupted with *LEU2*. pREd62 (Rhona Borts, Yeast genetics, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU) containing the *MSH3* gene disrupted with *LEU2* was digested to completion with the restriction enzyme *Aat*II (cf. Materials and Methods). Resulting transformants were then screened by Southern hybridization for the presence of the disrupted *MSH3* gene on chromosome III (data not shown).

(d) *MSH6*

The *MSH6* gene of *Saccharomyces cerevisiae* is a further gene that shares extensive homolgy with the bacterial *MutL* gene. MSH6p forms a heterodimer with MSH2p during the initiation of spontaneous base-base mismatch repair (Alani *et al.* 1996). Yeast strains mutant for *MSH6* exhibit a less profound increase in microsatellite instability as compared to *MSH3* mutant strains but have elevated levels of spontaneous mutations in comparison to *MSH3* mutant strains.

Disruption of MSH6

The *MSH6* gene of *Saccharomyces cerevisiae* YN94-1 was disrupted with *KanMX4*. This *MSH6* disruption plasmid (Rhona Borts, Yeast genetics, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU) was created by cloning a 4 kb PCR fragment containing the *MSH6* gene into the *Srf*I site of pPCR script (Sratagene). A *Pvu*II to *Eco*RV fragment containing the KanMX4 module (Wach *et al.* 1994) was then used to replace a *Sna*BI to *Spe*I fragment of the *MSH6* open reading frame

to create plasmid pSRC9. The *msh6::KanMX4* disruption cassette was released by digestion with restriction enzymes *Sph*I and *Bsp*EI and transformed into YN94-1 (cf. Materials and Methods). Resulting yeast transformants were then screened by Southern hybridization for the presence of the disrupted *MSH6* gene on chromosome IV.

Figure 15. Southern blot analysis of YN97-150 (*msh6::KanMX4*)

Southern blot analysis of *msh6* genomic DNA (Fig. 15A) and the corresponding disrupion scheme (cf. Fig. 15B). Lane (1) *Bam*HI digested control YN94-1 DNA. No bands are seen when probed with the 1553 bp KanMX DNA. Lane (2) *Bam*HI digestion of genomic DNA containing the disrupted *msh6* gene. A single band of 2481 bp is seen corresponding to the insertion of KanMX4.

(d) RTH1

The *RTH1* (*RAD27*) gene of *Saccharomyces cerevisiae* is one of several genes encoding a 5'→3' DNA exonuclease. Subsequent mutations arising in *rth1* strains are duplications resulting from a novel mutagenic process, and are not due to a defect in mismatch repair (Tishkoff *et al.* 1997). During replication of the lagging strand, DNA polymerase extends into the downstream "Okazaki" fragment and displaces it, resulting in a 5' "flap" structure that is normally removed by RTH1p. In the absence of this exonuclease, extensive strand displacement synthesis occurs resulting in the duplication of DNA sequences. Recently, it has been shown that expansion of CAG repeat tracts are frequent in yeast strains defective in Okazaki fragment maturation (Schweitzer *et al.* 1998). This result

supports the hypothesis that tract expansions that occur during passage of human disease allele results from excess DNA synthesis on the lagging strand of replication.

Disruption of *RTH1*

The *rth1::KanMX4* disruption strain was constructed using standard techniques. RTH1 was amplifyed by PCR using primers $_{-380}$ 5'-GTTAACAGTGACTTTCGGTGACAGATA-3' $_{-354}$ (numbers in subscript indicate positions relative to the start codon of *RTH1*) and 5'-GTTAACAAAGCTAGGTGTCGAAGG-3'$_{1424}$ (extra nucleotides were added to the 5'-end of this primer creating a HpaI site as underlined). These primers were made in-house at the 40 nmole scale.

Conditions for amplification were as follows: 50 ng YN94-1 genomic DNA template, 0.8 µl 25mM dNTPs (dATP, dCTP, dGTP, dTTP, Pharmacia) mix, 1µl each primer (200 pmole), 10 µl 10x Amplitaq® (Applied Biosystems) buffer, 2 units Amplitaq® enzyme made up to 100 µl with dist. H_2O and overlaid with paraffin oil. A Hybaid thermocycler was used. Programme: 92°C for 2 min for 1 cycle; then 92°C for 2 min, 50°C for 3 min, 72°C for 3 min. for 30 cycles. The reaction was then precipitated using 10 µl sodium acetate pH5.2 and 200 µl ethanol for 30 min on ice. The DNA was then centrifuged at 15000g for 20 min. The pellet was then washed in 75% ethanol and dried in vacuo before being resuspended in 30 µl H_2O.

The resulting PCR product was cloned into *Sma*I-digested pUC19 (Yannisch-Perron *et al.* 1985) creating plasmid pKLRTH1. A 1481 bp

*Sma*I/*Eco*RV-derived fragment containing the *KanMX4* cassette was removed from pFA6-KanMX4 (Wach *et al.* 1994) and cloned into the EcoRV site of pKLRTH1 (+66 relative to the *RTH1* start codon) creating disruption plasmid pKL*rth1*::*KanMX4*. The *rth1* disruption cassette was released by *Hpa*I digestion and then transformed into YN94-1. For selection of transformants geneticin (G418, Life Technologies) was added to the media at a final concentration of 200 μg/ml. Putative yeast disruptions were checked by Southern hybridization for the disrupted gene on chromosome XI and for the *rth1* temperature sensitive phenotype (Sommers *et al.* 1995).

Figure 16. Southern blot analysis of YN97-167 (*rth1::KanMX4*)

Southern blot analysis of *rth1* genomic DNA (Fig. 16A) and the corresponding disruption scheme (cf. Fig. 16B). Lane (1) *Eco*RI digested control YN94-1 DNA. One band is seen when genomic DNA is hybidized with the 316 bp PCR probe. This 3794 bp EcoRI fragment corresponds to the wild type *RTH1* gene on chromosome XI. Lane (2) *Eco*RI digestion of genomic DNA containing disrupted *rth1*. A shift in the 3794 bp band is seen to 2420 bp corresponding to the insertion of the KanMX4 cassette (this cassette contains an *Eco*RI restriction site at its 3'-end).

(e) ***POL30***

The *POL30* gene of *Saccharomyces cerevisiae* encodes PCNA (proliferating cell nuclear antigen), an essential component of the DNA replication machinery. PCNA exists *in vivo* as a homotrimer that acts as a

"sliding clamp" around the DNA helix thereby increasing the processitivity of DNA polymerases δ and ε. It has been shown that PCNA interacts strongly with the MSH2p/MSH3p heterodimer during mismatch repair (Johnson *et al.*,1996) Yeast strains carrying the *pol30-104* mutation show a dramatic increase in spontaneous mutation and tract instability. (cf. Table 1a). Plasmid pCH1577 (Amin & Holm, 1996) containing the *pol30-104* allele marked with the *LEU2* gene was digested to completion with *Sac*I. The reaction products were transformed directly into *S. cerevisiae* YN94-1. Putative disruptions were patched and first checked for the cold sensitive phenotype (inability to grow at 15 C) then by Southern hybridization analysis and finally by sequencing which revealed the expected nucleotide exchange of a C to a T at position 752 which results in the replacement of alanine by valine in the translated product.

Figure 17. Southern blot analysis of YN98-3 (*pol30-104::LEU2*)

Southern blot analysis of *pol30-104* genomic DNA (Fig.17A) and the corresponding disruption scheme (cf. Fig. 17B). Lane (1) *Eco*RV digested control YN94-1 genomic DNA. When hybridized with a *LEU2* probe two bands are seen (2861 and 961 bp corresponding to the wild type *LEU2* gene on chromosome III. Lane (2) *Eco*RV digestion of genomic DNA containing *pol30-104* allele. The *LEU2* probe hybridizes to four fragments, firstly the two wild type *LEU2* fragments as above and secondly two fragments (appearing as a doublet of 2067 and 2076 bp) corresponding to the *LEU2* marker inserted upstream of the POL30 locus

(f) Δmsh2 - deletion of *MSH2* in an α - type yeast strain

In general, it can be said that mutations affecting repeat stability are either dependent or independent of the MSH2p repair pathway. Mutations of *MSH2* together with one or more genes within this epistatic group show a mutator phenotype no greater than the single *MSH2* mutant alone. Mutations that are independent of this epistatic group together with the *msh2* allele show an additive or even a multiplicative effect with regard to the mutator phenotype.

We have deleted *MSH2* in an α-type yeast strain, which, when crossed with any a-type mutant yeast strain will allow the combination of the mutant and *msh2*. In this way, it can be ascertained whether the primary mutation is dependent or independent of the MSH2p pathway.

PCR-mediated disruption of *MSH2* by SFH

The SFH disruption cassette (Wach et al. 1994) was amplified from pFA6-KanMX4 by PCR using Amplitaq® (Applied Biosystems). Two primers were designed consisting of 40 bases of flanking sequence of *MSH2* and 20 bases complementary to the KanMX4 cassette.

The primers

SFH1:

5'-TTATCTGCTGACCTAACATCAAAATCCTCAGATTAAAAGTAG CTGAAGCTTCGTACGCTG -3'

SFH2:

5'-TATCTATCGATTCTCACTTAAGATGTCGTTGTAATATTAAATA GGCCACTAGTGGATCTG -3'

Conditions for the amplification were as follows: 50 ng pFA6-KanMX4, 0.8 µl 25 mM dNTPs mix, 1 µl each primer (200 pmole), 10 µl 10x Amplitaq® (Applied biosystems) buffer, 2 units Amplitaq® made up to 100 µl with dist. H_2O and overlaid with paraffin oil. For PCR programme see "disruption of RTH1". The PCR product was then precipitated using 10 µl 3 M sodium acetate pH 5.2 and 200 µl ethanol for 30 min on ice. The DNA was the centrifuged at 15,000 g for 20 min. The pellet was then washed in 75% ethanol and vacuum dried before being resuspended in 50 µl of dist. H_2O. Transformation of yeast was carried out according to the method of Gietz and Woods (1994). For selection, geneticin (G418)-supplemented medium was used. Putative yeast disruptions were checked by Direct Colony PCR (Pearson and McKee, 1991) using the following primers.

Checking primer 1:

$^{-269}$ 5'- AGCACTCCGTATAAACAAAG -3' $^{-250}$

Checking primer 2:

^{3111} 5'- TAGTGACAGTGGAATAAAGG -3' ^{3130}

(Numbers in superscript indicate positions relative to the MSH2 start codon).

On checking by PCR a wild type *MSH2* strain gives a product of approximately 3399 bp whereas strains in which *MSH2* has been disrupted by the KanMX4 cassette yield a product of approximately 2016 bp.

Figure 14. Southern blot analysis of YN97-11 (*mlh1::LEU2*)

Lane (1) *Eco*RV digested control YN94-1 DNA. Two bands are seen

(approximately 1000 and 3000 bp) corresponding to the wild type *LEU2* gene on chromosome III. Lane (2) *Eco*RV digests of yeast genomic DNA containing disrupted *MLH1*. The gel was probed with a 2200 bp *LEU2* fragment. Three bands are seen in the disruptant, two corresponding to the wild type *LEU2* on chromosome III and the other larger band corresponding to the *LEU2* gene used to disrupt *MLH1* on chromosome XIII (approximately 5800 bp).

METHODS

1. ***E. COLI* TRANSFORMATION USING THE "CALCIUM CHLORIDE" METHOD**

(Mandel & Higa 1970)

Solutions:

0.1 M $MgCl_2$

0.1 M $CaCl_2$

Protocol:

An overnight culture was diluted (1:100) in 50ml of LB medium and grown at 37°C with shaking until OD_{650} ≈ 0.3-0.35 (2-2.5 hr approx.). The culture was divided into 2x25ml tubes and incubated on ice for 20 minutes. Tubes were spun at 4°C for 3 minutes at 3500rpm in a centrifuge (Sigma 4k10). The supernatant was discarded and 20ml 0.1 M $MgCl_2$ was added, vortexed and then immediately spun. 10ml of sterile ice-cold $CaCl_2$ was added and left on ice for 10 minutes. Tubes were spun

and pellet was resuspended in 2ml 0.1 M $CaCl_2$. Cells were then frozen at -70°C or used directly for transformation. For transformation, 100µl of cell suspension was added to a pre-cooled reaction tube together with 4µl (1-10ng) of transforming DNA. This was then left on ice for 30 minutes. Cells were then heat-shocked for 45 seconds at 42°C and then incubated on ice for 4 minutes. 1ml of LB medium was then added and the cells were incubated at 37°C for 1 hour. This was then plated directly onto selective media LB plates and incubated overnight at 37°C.

2. **YEAST TRANSFORMATION USING THE "PLATE" METHOD**

(Elble, 1992)

Solutions:

PLATE: 90ml 45% PEG 4000
10ml 1M Li-acetate
1ml 1 M Tris-HCI, pH 7.5
0.2ml 0.5 M EDTA

Protocol:

The yeast strain was streaked out for single colonies onto YEPD agar and incubated overnight at 30°C. A single colony was picked and grown overnight in 10ml YEPD at 30°C with shaking. 0.5ml of the culture was spun down in a microcentrifuge (Sigma 112) and the supernatant was decanted by inversion. 10µl of carrier DNA (100µg) and 1µg transforming DNA was then added (no transforming DNA was added for

control) and vortexed. 0.5ml PLATE was then added, vortexed and incubated overnight at room temperature. Cells were then pelleted, washed and resuspended in 125µl glass dist. H_2O. This was then plated directly onto selective media and incubated at 30°C for 3 days.

3. **RAPID ALKALINE EXTRACTION PROCEDURE FOR SCREENING RECOMBINANT PLASMID DNA (MINISCREENS)**

(Birnboim & Doly. 1979)

Solutions:

Resuspension buffer: 50 mM Tris-HCl
10 mM EDTA, pH 8.0

Lysis buffer: 200 mM NaOH
1% SDS

Neutralization Buffer: 3 M Na acetate, pH 5.5

Isopropanol

Phenol/Chloroform/isoamyl alcohol (25:24:1)

Chloroform/isoamyl alcohol (24:1)

Na-acetate 3 M

Ethanol 100% and 70%

Protocol:

1.5ml of an overnight culture was spun in a bench centrifuge (Sigma 112) for 3 minutes and the supernatant decanted. 200µl of resuspension buffer

was then added and vortexed. 200µl of lysis buffer was then added and the tubes inverted six times 200µl of neutralization buffer was added immediately and the tubes inverted six times. This was then spun for 15 minutes in a bench centrifuge and the supernatant transferred to a new tube. DNA was then precipitated with 0.6 vol. isopropanol, spun and the pellet washed with 70% ethanol. This was then dried *in vacuo* and resuspended in 50µl glass dist. H_2O.

4. PREPARATIONS OF YEAST GENOMIC DNA

(Sherman *et al.* 1982)

Solutions:

Sorbitol/Tris/EDTA: 1.2 M sorbitol
10 mM EDTA
0.1 M Tris-HCl, pH 8.0
Dithiothreitol (DTT) 1 M

Zymolyase 20T, 5mg/ml (ICN Biomedicals Inc., Oxford, UK)

RNase A, 10mg/ml (Sigma, Dorset, UK) - boiled for 10 minutes, then snap cooled.

Pronase E, 10mg/ml (Sigma, Dorset, UK)

SDS 10%

Phenol/chloroform/isoamyl alcohol (25:24:1)

Chloroform/isoamyl alcohol (24:1)

Na-acetate 3 M

Isopropanol

Ethanol 75%

Protocol:

An individual yeast colony was used to inoculate 10ml SCD-U and grown up overnight at 30°C with shaking. The culture was then spun at 3000rpm for 3 minutes and resuspended in 5ml sorbitol/Tris/EDTA, 10µl DTT was added and the culture incubated at 30°C for 30 minutes. The cells were pelleted as before and resuspended in 0.5ml sorbitol/Tris/EDTA. 10µ Zymolyase 20T and 10µl RNase A were added and incubated at 30°C. 10µl pronase E and 50µl SDS were added and incubated at 37°C for 2 hours. One phenol/chloroform/isoamyl alcohol and three chloroform/isoamyl alcohol extractions were then carried out. 0.1 vol. 3 M Na-acetate and 0.6 vol. isopropanol were then used to precipitate the DNA which was pelleted and washed with 70% ethanol. This was then air dried and dissolved in 100µl glass dist. H_2O.

5. **RESTRICTION DIGESTS**

(Maniatis *et al.* 1989)

Solutions:

Restriction Buffers (working concentrations) (Life Technologies Inc., Paisley, UK):

REact™1	50 mM Tris-HCl (pH 8.0)
	10 mM $MgCl_2$
REact™2	50 mM Tris-HCl (pH 8.0)
	10 mM $MgCl_2$

50 mM NaCl

REact™3 50 mM Tris-HCl (pH 8.0)
10 mM $MgCl_2$
100 mM NaCl

REact™6 50 mM Tris-HCl (pH 7.4)
6 mM $MgCl_2$
50 mM KCl
50 mM NaCl

Protocol:

For purification of DNA fragments, restriction digests were performed in a total volume of 30µl. 20µl of DNA (0.2-1µg) was digested with 2µl of enzyme, 3µl of 10x restriction buffer and 5µl glass dist. H_2O. The reaction was incubated according to the manufacturer's instructions for a minimum of 3 hours. For restriction analysis, digests were performed in a total volume of 10µl. 1µl of DNA was digested with 1µl of enzyme, 1µl of the appropriate buffer and 7µl glass dist. H_2O. This was then incubated according to the manufacturer's instructions for 1 hour. Enzymes were then inactivated by heating at 60°C for 10 minutes.

---

6. **RECOVERY OF DNA FROM LOW-MELTING-TEMPERATURE AGAROSE**

(Weislander, 1979)

Solutions:

Tris-borate (1xTBE) buffer: 0.089 M Tris-base
0.089 M boric acid
0.002 M EDTA pH 8

Ethidium bormide (working conc. 0.5 mg/ml)
Tris-HCI 20 mM pH 8
EDTA 1 mM pH 8
Phenol/Chloroform/isoamyl alcohol (25:24:1)
Chloroform/isoamyl alcohol (24:1)
Isopropanol
Ethanol 70%

Low-melting-temperature agarose (Life Technologies Inc., Paisley, UK).

Protocol:

0.25g of low-melting agarose was dissolved in 25ml 1xTBE and heated to 70°C. The gel was then cast and allowed to set at 4°C. DNA samples were then loaded into the gel and electrophoresis was carried out at 60v. The gel was then stained for 30 minutes in ethidium bromide and the band of interest was localised using a long-wave-length (300-360 mn) uV lamp to minimize damage to the DNA. The band was then cut out of the gel with a flamed scalpel and added to 5 vol. 20 mM Tris-HCI and 1 mM EDTA. This was then heated at 65°C to melt the gel. DNA was then extracted using standard procedures and redissolved in 50µl glass dist. H_2O.

7. DNA LIGATIONS

(Maniatis *et al.* 1983, Pheiffer & Zimmerman, 1983)

Solutions:

T4 DNA ligase (Promega-Biotec, Madison, USA)

1x ligase buffer (Promega-Biotec, USA): 50 mM Tris-HCl, pH 7.8
10 mM $MgCl_2$
10 mM dithiothreitol
1 mM ATP
25 mg/ml bovine serum albumin

Protocol:

Ligations were carried out in a total volume of 10µl. Approximately 0.1µg of vector DNA and the appropriate amount of insert DNA was added to a sterile microfuge tube. Reaction was made up to 7.5µl with glass dist. H_2O. For cohesive end ligation, the reaction was warmed to 65°C for 5 minutes, 37°C for five minutes, room temperature for five minutes, then 4°C for five minutes. 1µl ligase buffer and 1µl T4 ligase was added and the reaction was then incubated overnight at 20°C for cohesive ends, or at 10°C for the ligation of blunt-ended DNA. 1-4µl of the ligation mixture could then be used for transformation of *E. coli*.

8. SOUTHERN HYBRIDIZATION

(Southern, 1975, ECL™ Technical Manual, Amersham, UK, 1951)

This direct nucleic acid labelling and detection system is based on enhanced chemiluminescence. The system involves directly labelling probe DNA with the enzyme horseradish peroxidase, achieved by denaturing the probe so that it is in single stranded form. Subsequent addition of glutaraldehyde causes the formation of chemical cross-links so that the probe is covalently labelled with enzyme. Reduction of hydrogen peroxide by the enzyme is then coupled to a light producing reaction involving luminol, which on oxidation produces blue light which can be detected on light sensitive film.

Solutions:

Tris-borate TBE buffer: (cf. earlier)

Agarose (Life technologies Inc., Paisley, UK)

0.25 M HCl

Denaturation solution:	1.5 M NaCL 0.5 NaOH
Neutralization solution:	1.5 M NaCl 0.5 M Tris-HCl pH 7.5
20xSSC	3 M NaCl 0.3 M Tri-sodium citrate pH 7.0
Hybridization buffer 0.2 M:	(Amersham Life Sciences, ECL Gold hybridization buffer, with 5% (w/v) blocking agent and 2.2 M NaCl)

SDS 1%

Hybond-N (Life Science, Amersham, UK)

ECL Nucleic Acid Labelling and Detection Kit (Life Science, Amersham, UK).

Protocol:

Pre-digested genomic DNA was run on a 1% agarose gel in 1xTBE. The gel was then immersed in 0.25 M HCI and agitated for 7 minutes. This was then rinsed twice with dist. H_2O and immersed in denaturation solution for 30 minutes. After rinsing with deionised H_2O the gel was immersed in neutralization solution for 30 minutes. The DNA was then transferred to Hybond-N with 20xSSC using a vacuum blotter (Appligen, Durham, UK) at 60 mbar for 1 hour. DNA was cross-linked onto the membrane in a Stratalinker 2400 (Stratagene, Cambridge, UK), using the autocrosslink mode (120,000 µjoules/30 sec). The blot was then placed in a glass hybridization tube (80 x 200 mm, Techne, Cambridge, UK), and 25ml hybridization buffer was added. The blot was then pre-hybridized for 15 minutes at 42°C.

9. PREPARATION OF PROBE

The DNA fragment to be labelled was diluted to a concentration of 10ng/µl in glass dist. H_2O. 100ng (10µl) was then denatured by boiling for 5 minutes. The DNA was immediately cooled on ice for 5 minutes and briefly spun in a microcentrifuge. 10µl of ECL DNA labelling reagent was added to the cooled DNA and mixed. 10µl of ECL glutaraldehyde was added and mixed. This was then incubated at 37°C for 10 minutes. The probe was then added to the hybridization solution at a final concentration of 10ng/ml.

The blot was hybridized overnight in a Techne-oven (Hybridizer HB-1D, Techne, Cambridge, UK) at 42°C and then washed at 65°C with 2x SSC and 1% SDS. A second wash was carried out with 0.5x SSC and 0.1% SDS. The blot was then exposed to luminescence detection film (Hyperfilm-ECL, Amersham, UK) for various lengths of time.

10. **DIRECT COLONY PCR**

(Pearson & McKee, 1992)

Solutions:

Glass dist. H_2O

Taq polymerase buffer 10x

(Perkin Elmer, Cheshire, UK) 500 mM KCl

100 mM Tris-HCl pH 9

1% Triton X-100

Taq polymerase (Perkin Elmer, Cheshire, UK)

Primer: *Lac*Z Reverse 1:50 dilution

Primer: YEp Forward 1:50 dilution

dNTPs, 25 mM

Sterile Oil

PCR purification resin (Promega Biotech, Madison, USA)

Protocol:

Single yeast colonies were picked with a sterile tip and mixed with 10µl glass dist. H_2O in a 0.5ml reaction tube and placed on ice. A mix

containing 100µl polymerase buffer, 780µl H_2O, 8µl dNPTs, 4µl of each primer and 5µl *Taq* polymerase was then made up and vortexed briefly. 90µl of this mix was added to each reaction tube and sterile oil was added to prevent evaporation.

PCR Programme (Thermal reactor, Hybaid, Middlesex, UK):
92°C - 2 min
1 cycle

92°C - 2 min
45°C - 3 min
72°C - 2 min
30 cycles

The majority of the oil was pipetted off. The PCR product was purified using PCR purification resin according to the manufacturer's instructions and redissolved in 100µl glass dist. H_2O.

The reaction products were identified by running 15µl of the assay on a 2% agarose gel in 1xTBE.

11. YEAST β-GALACTOSIDASE ASSAY

(Miller, 1972, Guarente, 1983)

Solutions:

Phosphate buffer (KPP) 0.1M: 1 M KH_2PO4
1 M K_2HPO4

Mix both until pH 6.5, then dilute 1:10

Z buffer: 60 mM Na_2HPO_4
40 mM $NaH_2PO_4.H_2O$
10 mM KCl
1 mM $MgSO_4$

Na_2CO_3 1 M

β-mercaptoethanol (Millipore, Bedford, UK)

0-Nitrophenyl β-D-Galacto-pyranoside (ONPG) (Sigma, Dorset, UK): 4mg/ml

0.10-0.11 mm acid-washed glass beads (B. Braun Melsungen, Germany)

Protocol:

10 ml overnight yeast culture in SCD-U was spun for 3 minutes at 5000rpm (Sigma 5k10). The supernatant was decanted and the pellet washed twice with KPP buffer. The pellet was then frozen at -20°C for 3 hours to "crack" the cells. The pellet was placed on ice and the same volume of glass beads was added, this was then vortexed at maximum speed for 1 minute. 1.2ml of Z buffer was then added and vortexed. This was then spun in a pre-cooled centrifuge (4°C) for 5 minutes at 3000rpm and the extract decanted into fresh chilled tubes. For the assay 25µl of extract was then added to 1ml Z buffer + β-mercaptoethanol (175µl β-mercaptoethanol in 50 ml Z buffer) and warmed to 30°C (control; 1 ml Z buffer / β-mercaptoethanol + 25µl Z buffer). 200µl ONPG soln was added and the time measured for a yellow colour to appear. The reaction was then stopped with 500µl M Na_2CO_3 soln and the OD_{420} measured

(Ultraspec 2000, Pharmacia Biotech, St. Albans, UK).

12. DETERMINATION OF VOLUME ACTIVITY (U/ML) IN CELL EXTRACTS

Volume activity was then calculated for each sample using the following equation:

$$\text{VOLUME ACTIVITY} = 1000 \times \frac{V_T}{4.5 \times V_E \times dt} \times OD_{420}$$

Where V_E = extract volume (25µl)

V_T = total reaction volume (1725µl)

dt = time (min)

Unit Definition (U) = one unit (U) of β-galactosidase hydrolyses 1 nMol ONPG per minute under the above conditions.

13. DETERMINATION OF TOTAL PROTEIN IN CELL EXTRACTS

To determine specific β-glactosidase activity, total cell extract protein must be quantified.

Solutions:

Bradford solution (Bio-Rad, Munich)

Protoc l:

25µl cell extract was made up to 800µl with glass dist. H_2O. 200µl of Bradford solution was then added, vortexed and incubated at room temperature for 10 minutes. OD_{595} was then measured (Ultraspec 2000, Pharmacia Biotech, St. Albans, UK) and protein concentration (mg/ml) was determined using a bovine serum albumin calibration curve.

Determination of specific activity (U/mg) in cell extracts

Specific activity of the cell extracts could then be calculated using the following equation:

$$\text{SPECIFIC ACTIVITY} = 1000 \times \frac{V_T}{4.5 \times V_E \times dt \times Cp} \times OD_{420}$$

Where V_E = extract volume (25µl)

V_T = total Reaction volume (1725µl)

Cp = protein Concentration (mg/ml)

dt = time (min)

Unit Definition (U) = one unit (U) of β-galactosidase hydrolyses 1 nMol of ONPG per minute under the above conditions.

14. **GEL PURIFICATION OF OLIGONUCLEOTIDES**

Solutions:

4% polyacrylamide gel:	2.5ml	glycerol
	2.5ml	acrylamide/N,N'-Methylenebisacrylamide 40% (w/v) (Millipore, Bedford, UK)
	5ml	5xTBE
	15ml	H_2O
	200µl	10% (w/v) ammonium persulphate (APS)
	12µl	N,N,N',N'-Tetra-methyl-ethylenediamine (TEMED), (Bio Rad, Herts UK)

Protocol:

50µg of each oligo was run on a 4% acrylamide gel in 1xTBE at 100 volts. Single stranded DNA bands were shadowed by UV using a fluorescent TLC plate and cut out. DNA was eluted from gel fragments into 200µl glass dist. H_2O by shaking overnight and precipitated using standard procedures.

15. <u>DNA SEQUENCING</u>

(Sanger *et al.*, 1977)

Protocol:

Nucleotide sequences were determined by automated DNA sequencing based on the chain-termination method using the ABI 373A sequencer (Applied Biosystems, Foster City, California, USA). Double stranded DNA was sequenced using the 'Taq DNA polymerase dideoxy terminator cycle sequencing kit' (Applied Biosystems), with primers; *lacZ* Reverse

and YEp Forward (cf. Materials and Methods).

EXAMPLE 1:

Construction of preferred yeast assay plasmids

pKa3-9(32-1), pKaCEN(32-1) & pKaINT(32-1)

The preferred system utilises a unique group of yeast vectors containing the bacterial *lacZ* gene (minus the promoter and first 7 codons) fused to the first 29 codons and promoter region of the yeast gene *PRS3* (5-phospho-ribosyl-1(α)-pyrophosphate synthetase) (Carter *et al.* 1994). This functional gene fusion was preferably knocked out-of-frame (-1 reading-frame) by the insertion of a poly $d(AC)_{16}$ tract at the *Eco*RI site downstream of the *PRS3* promoter and within the coding region initiated from the ATG of *PRS3*. The resulting *PRS3'/lacZ* gene fusion containing this out-of-frame insertion was inserted into each of three yeast vectors: a high copy vector, a low copy vector and a single copy integrated into the yeast genome.

1(i) **pKa3-9(32-1)** (high copy construct)

The initial *PRS3'/lacZ* fusion was constructed by the insertion into the multiple cloning region of YEp356R (Fig. 2, Materials and Methods) (Myers, A.M., *et al.* 1986) of a 371 bp DNA fragment comprising the promoter region and regulatory elements of the yeast gene *PRS3*. This 371 bp *HpaI/ClaI* fragment (-284 - +85 relative to the *PRS3* start codon) was rendered blunt and the resulting fragment ligated into the unique *SmaI* site of YEp356R, creating plasmid pSS3-9. Oligonucleotides 1 and 2 (cf. Materials and Methods) containing 16 repetitive AC or TG units were

annealed and cloned into the *Eco*RI site situated within the multiple cloning region of pSS3-9 (50μg of each oligonucleotide was purified on a 4% acrylamide gel and redissolved in glass dist. H_2O. 200ng of each olinucleotide pair were then annealed in glass dist. H_2O by heating to 90°C and allowing to cool slowly to room temp. pSS3-9 was cut to completion with *Eco*RI in buffer REact™3 and subsequent ligation was carried out with 100-fold excess of insert to vector at 12°C (cf. Materials and Methods)). Ligations were transformed directly into *S. cerevisiae* YN94-1 using the "plate" method (cf. Materials and Methods) and resulting colonies were screened for the presence of insert by "direct colony PCR" (cf. Materials and Methods). Insertion of this repetitive tract creating assay plasmid pKa3-9(32-1) resulted in shifting the coding region of the *PRS3'/lacZ* fusion to the -1 reading frame. Also contained within plasmid pKa3-9(32-1) are sequences from the yeast 2 micron origin of replication which allows this construct to exist episomally at a copy number of 50-150 per cell. pKa3-9(32-1) also contains the *Col*EI origin permitting replication in *Escherichia coli*. Selection for the assay plasmid is allowed by the incorporation of the ampicillin resistance gene of *E. coli* and the *URA3* gene of yeast respectively (see. Fig. 5).

1(ii) **pKaCEN(32-1)** (low copy construct)

Construction of this low copy centromere-based plasmid involved removal of the *PRS3'/lacZ* fusion containing the poly $d(AC/TG)_{16}$ tract from pKa3-9(32-1) on a 3680 bp *Bam*HI/*Nsi*I fragment (cf. Fig. 5). Initially pKa3-9(32-1) was digested to completion with *Bam*HI and *Nsi*I in restriction buffer REact™3 (Materials and Methods). KCI was added to adjust the salt concentration to that of REact™6 and digestion with Scal was then

carried out to eliminate incorporation of the smaller *Bam*HI/*Nsi*I fragment. The resulting 3680 bp fragment containing the *PRS3-'lacZ* fusion was gel purified (cf. Materials and Methods). pRS416 was also digested to completion with *Nsi*L and *BamH*I in buffer REact™3 with the aim of removing the existing *lacZ* region on a 923 bp fragment. The resulting 3975 bp vector band containing the ARS and centromere sequences was gel purified and ligated to the 3680 bp fragment of pKa3-9(32-1) containing the *PRS3/LacZ* fusion. In this way the 7655 bp plasmid - pKaCEN(32-1) (Fig. 6) was generated. The ligation mixture was then transformed into *E. coli* DH5α (cf. Materials and Methods) and placed onto LB + ampicillin plates. DNA from resulting colonies was purified using the rapid alkaline extraction procedure (cf. Materials and Methods) and screened by restriction analysis. The low copy centromere-based plasmid exists episomally at levels of 1-5 copies per cell (see Fig. 6).

1(iii) **pKaINT(32-1)** (single copy integrative construct)

Construction of pKaINT(32-1) involved removal of the *PRS3'/lacZ* fusion containing the poly $d(AC)_{16}$ tract from pKa3-9(32-1) on a *Bam*HI/*Nsi*I fragment. This fragment was subsequently cloned into *Bam*HI/*Nsi*I digested integrative plasmid YIp352 (cf. Fig. 4, Materials and Methods (Myers *et al.* 1986)). Before pKaINT(32-1) could be integrated into the yeast genome it was linearised by restriction at its unique *Nco*I site located within the *URA3* gene. Integration then occurred by homologous recombination between the *URA3* gene of pKaINT(32-1) and the genomic *URA3* locus of *S. cerevisiae* YN94-1 (Fig. 8). The resulting *URA3* strain was checked by Southern hybridization to confirm the insertion of the integrative plasmid (cf. Materials and Methods). (See Figs. 7, 8, 9 and

10).

2. **Southern blot analysis of integrated vector**

Southern hybridization was used to confirm the integration of pKaINT(32-1) into the yeast genome (cf. Materials and Methods). A 423 bp *Eco*RI derived restriction fragment from pKaINT(32-1) containing the yeast *PRS3* promoter region was used to probe YN94-1 genomic digests (Fig. 7). In 5 clones transformed with pKaINT(32-1), probing *Bam*HI/*Nsi*I-restricted yeast genomic DNA revealed two bands; firstly, a 2719 bp fragment arising from the wild type *PRS3* promoter region on chromosome VIII, and secondly a 3680 bp fragment corresponding to the engineering *PRS3* promoter of pKaINT(32-1), integrated at the *URA3* locus on chromosome V. Therefore the plasmid has been stably integrated into the genome (Figs. 8, 9 and 10).

From the Southern blot (Fig. 10) we see that lane 6 gives the same banding pattern as the control YN94-1 in lane 1. This pattern may be due to recombination between the repeated *URA3* genes on chromosome V. In this way the integrated vector is lost and the *URA3* mutation reverts to wild type, hence growth on selective media and loss of the 3680 bp band on the blot.

EXAMPLE 2:

Sensitive screening system in *S. cerevisiae* utilising the luciferase reporter gene

Here we describe the use of an alternative reporter gene in a screening in

S. cerevisiae, utilising the light emitting properties of the luciferase gene product. The development of this system involves simple sub-cloning steps with a fragment containing the luciferase gene derived from "pGL3-Promter vector" (Promega-Biotec) (cf. Fig. 15). This fragment is fused to our PRS3'/$(AC)_{16}$ tract using the three assay plasmids described above which have had the *lacZ* reporter gene previously removed. DNA stability is then monitored by alterations in levels of measured light emittance using a Lumat luminometer LB9501 (Berthold, UK) with the luciferin substrate obtained from Promega Biotec.

2(i) **Further example of a suitable reporter gene:**

The *URA3* (orotidine-5'-phosphate decarboxylase) gene from *Saccharomyces cerevisiae*

Isolated and sequenced by Mark Rose, Paula Grisafi and David Botstein. (1984); *Structure and function of the yeast URA3 gene: expression in Escherichia coli*. *Gene*, **29**, 113-124.

SGD Name: URA3/YEL021W (http://genome-www.stanford.edu/)

Cells in which the *URA3* gene is expressed can be selected against by growth on plates containing 5-fluoro-orotate (5-FOA).

EXAMPLE 3:

Preferred human assay vectors

Screening system utilising the β-galatosidase reporter gene for use in

human cell lines

This screening system involves the construction of a similar plasmid to that used in the yeast system. In this case a fragment containing the luciferase gene is removed from pGL3-Promter vector (Fig. 15) and replaced with a fragment derived from one of the three assay plasmids described in Example 1. This fragment derived from the assay plasmid contains the poly$(AC)_{16}$ tract fused to the β-galactosidase reporter gene minus the yeast promoter and start codon. Cloning of this fragment into pGL3-Promter vector results in an SV40/*lacZ* fusion containing a poly$(AC)_{16}$ tract in the open reading frame. Replication in human cell lines is allowed through the incorporation of the EBV (Epstein-Barr virus) origin of replication derived from plasmid pDR2 (cf. Fig. 16). Selection in human cultured cells is by incorporation of a fragment containing the hygromycin resistance gene also derived from plasmid pDR2 (cf. Fig. 16).

Rather than yeast cells as described in the previous Examples, in another embodiment human cell lines are utilized.

Preferably the cells are derived from tumours of colorectal cancer patients or other tumours which exhibit instability in repetitive (microsatellite) DNA sequences. Alternatively, human cell lines can be engineered to contain mutations in genes implicated in mismatch repair pathways, such as hMSH2 and/or hMLH1, using standard mutagenesis techniques within the knowledge of a skilled person.

The following vectors are used to transform human cell lines in the same way that the vectors of Examples 1 and 2 are used to transform yeast cells

according to preferred embodiments of the assay according to the first aspect of the invention.

EXAMPLE 4:

Assay for testing the carcinogenic properties of a test substance

Using the test system of the invention the background frequency of microsatellite instability associated with wild type (mismatch repair-competent) and strains defective in DNA mismatch repair has been assayed. As shown in Table 1 the yeast strains defective in mismatch repair show extensive DNA instability. This situation can be compared to the DNA instability observed in cell lines derived from colorectal carcinomas.

Table 1: Frequency of alternation in lengths of poly(AC) tracts in wild type yeast strains and DNA mismatch repair mutants.

Strain	Relevant genotype	Tract location	Frequency of tract alteration	Frequency relative to wild type
YN96-57	wild type	pKa3-9(32-1)	1.35×10^{-4}	1
YN97-25	*msh2*	pKa3-9(32-1)	6.80×10^{-2}	504
YN97-28	*mlh1*	pKa3-9(32-1)	1.14×10^{-1}	843
YN97-31	*msh3*	pKa3-9(32-1	6.06×10^{-3}	45
YN97-22	wild type	pKaCEN(32-1)	1.00×10^{-5}	1

YN97-26	*msh2*	pKaCEN(32-1)	$5.90x10^{-3}$	590
YN97-29	*mlh1*	pKaCEN(32-1)	$6.14x10^{-3}$	614
YN97-32	*msh3*	pKaCEN(32-1)	$1.40x10^{-3}$	139
YN97-23	wild type	*URA3* locus,	$<5.0x10^{-6}$	1
YN97-27	*msh2*	chr. V	$1.59x10^{-3}$	318
YN97-30	*mlh1*	*URA3* locus,	$1.82x10^{-3}$	363
YN97-33	*msh3*	chr. V	$8.45x10^{-5}$	9
		URA3 locus,		
		chr. V		

Basal frequency of microsatellite instability of the various genetic backgrounds within the yeast assay system

Using each type of assay construct frequencies of microsatellite instability were measured in relation to wild type yeast and disruptions as described in Table 1a. Slippage events responsible for the white to blue colony transition were determined by sequencing.

(I) Wild type

As shown in Table 1a the wild-type yeast strain for mismatch repair - YN94-1 - shows a low background white to blue frame-shifting frequency when transformed with constructs described above in the context of high copy 2 micron-based yeast origin of replication. Frequency of instability increases relative to the length of the repetitive tract (obviously a longer tract is more likely to undergo mutational change than a shorter one). Frequency of instabilities are approximately 10-fold lower in constructs in

the context of the centromere-based vectors compared to the high copy constructs and approximately 100-fold lower in the single copy integrative assay constructs. For the high copy construct containing the $(C_{1-3}A)_n$ telomeric tract the frequency of alteration was approximately $3x10^{-5}$ suggesting that telomeric repeats are much more stable than poly (AC) tracts. In this way, the overall sensitivity of the assay system can be altered through a choice of plasmid copy number and repetitive tract length. Little difference in slippage-frequency was seen between the constructs in the -1 reading-frame compared to those in the +1 reading frame when measured in the wild type strain, indicating no bias towards insertions or deletions. The sequence of the d(AC) tract in plasmids rescued from blue colonies was determined as described. As an example, for assay constructs containing the poly $d(AC)_{16}$ tract (-1 reading-frame) the only alteration detected was the loss of one (AC) pair which resulted in a poly $d(AC)_{15}$ tract. This seemed to be irrespective of plasmid copy number or tract location. Using this strain (cf. Table 1a) 14 of the 20 high copy plasmids in the -1 reading-frame analysed had altered tract lengths. The 6 blue colonies with unaltered tracts presumably had a mutation elsewhere in the *LacZ* gene. Alternatively as this is a high copy construct the subpopulation of altered tracts responsible for the white to blue transition may not have been rescued. This latter theory was strengthened by the finding that 10 out of the 10 low copy centromere-based (-1) plasmids rescued (cf. Table 1a) contained altered tract lengths. For blue yeast colonies containing chromosomally integrated $poly(AC)_n$ tracts a PCR was carried out as described. Four PCR products corresponding to $(AC)_n$ tracts integrated at the *URA3* locus on chromosome V were analysed. Sequencing showed that each repetitive tract had lost a single (AC) pair. The loss of this (AC) pair resulted in a

shift of the coding region from the -1 reading frame back into the correct reading frame thereby giving rise to a functional gene product. For assay constructs containing the poly $d(AC)_{12}A$ tract (+1 reading-frame) the most frequent alteration detected was the gain of one (AC) pair which resulted in a poly $d(AC)_{13}A$ tract. Again, this was irrespective of plasmid copy number or tract location. In the wild type strain YN94-1 (cf. Table 1) 9 of the 12 high copy plasmids analysed contained altered tract lengths. 8 of these contained one extra (AC) pair, the 9th contained an extra 4 (AC) pairs giving a poly $d(AC)_{16}A$ tract. The gain of a single (AC) pair or in one case 4 (AC) pairs resulted in the shifting of the coding region of the *PRS3'/lacZ* fusion from the +1 reading frame back into the correct reading frame giving rise to a functional gene product.

(II) *msh2*

A 100-600 fold increase in the frequency of microsatellite instability was measured in yeast strain YN97-10 carrying a disrupted *msh2* allele (Table 1a). Difficulty was encountered in determining frequencies of slippage in this strain carrying high copy plasmids due to the number of mutational events observed. The number of slippage events indicated by the white to blue colour transition was found to be higher in studies utilising the assay constructs in the -1 reading-frame compared to studies utilising assay constructs in the +1 reading-frame, indicating a bias for deletions over insertions. This finding is consistent with the previous results of Strand *et al.* (1993) who also found a bias towards deletions for this strain. The values clearly illustrate MSH2p as being a key component in the mismatch repair pathway with a strategic and unique role in the identification and correction of insertion/deletion loop structures in the genetic material.

(III) *mlh1*

Mutation in the yeast mismatch repair gene MLH1 (Table 1a) also has a profound effect on the rate of microsatellite instability within all the assay constructs. The degree of instability is in the same order of magnitude to that of *msh2* mutants (approx. 200-800 fold, cf. Table 2). Again frequencies of frame-slippage are higher in studies utilising the assay constructs in the -1 reading-frame. Although the exact function of MLH1p is unknown these results again illustrate the importance of this gene product in the correction of mispaired loop structures.

(IV) *msh3* and *msh6*

Mutation in the yeast genes *MSH3* and *MSH6* have less profound effects on the rate of microsatellite instability within this yeast system as compared to mutations in *MSH2* and *MLH1* (Table 1a). Mutation in the *MSH3* gene causes an approx. 30-fold increase in frame-slippage compared to wild type. An approximate 2-fold increase in white to blue transition was seen in studies using the high copy (-1) construct compared to the (+1) construct and a 4-fold increase was seen in the levels of slippage using the centromere-based (-1) construct compared to the centromere-based (+1) version, again indicating a bias towards deletions rather than insertions in this strain. This finding is also consistent with the previous findings of Strand *et al.* (1995) who found a bias towards deletions for the strain deleted in *MSH3*. Mutation in the *MSH6* gene causes an 8-12 fold increase in insertion/deletion events indicating the secondary role this mutS homologue plays in the pathway.

(V) *rth1*

We have shown that strains mutated for the gene encoding this 5'→3' exonuclease RTH1p, cause a 20-260 fold increase in the frequency of frame-slippage compared to wild type. We have found that this is also dependent on the reading-frame of the assay construct used. Previous data (Johnson *et al.* 1995) have shown that mutations in *RTH1* cause up to a 280-fold increase in instability of simple repeats compared to wild-type strains. Secondly, tract changes involving single repeats are exclusively insertions in the *rth1* strain consistent with the theory of displacement synthesis. Similar to these results we have found a 260-fold increase in instability for centromere-based assay constructs in the +1 reading-frame and a 20-fold increase for assay constructs in the -1 reading-frame (Table 1). On sequencing of these tracts we found that in assay constructs containing tracts in the +1 reading-frame all alterations were due to the insertion of one single repeat unit whereas all the alterations in tracts originating in the -1 reading-frame showed the loss of one (AC) pair or in one case the loss of four (AC) pairs. No insertions were found. These results suggest that as well as the prevention of displacement synthesis leading to the duplication of DNA sequences, RTH1p may also play a role in the prevention of loops occurring in the template strand.

(VI) *pol30*

The *pol30-104* mutation in PCNA causes an approximately 20-100-fold increase in the frequency of tract alteration as compared to wild type. The effect of *pol30-104* on tract alterations, however, is not quite as severe as

that of the null mutations in the mismatch repair genes (cf Table 1a). Epistatic analyses of *pol30-104* with null mutations in mismatch repair genes *MSH2*, *MLH1* and *PMS1* have shown that rates of tract instability were the same in double mutants of *pol30-104* with null mutations in mismatch repair genes and in single mismatch repair mutants (Johnson *et al.* 1996). Therefore, it can be said that hypermutability in this mutant results from a defect in mismatch repair.

4(i) Determination and verification of frame-slippage events

When transformed into yeast the preferred three assay plasmids produce white colonies on medium containing a chromogenic substrate for β-galactosidase and no β-galactosidase activity can be measured (cf. Materials and Methods). Frame-slippage within the repetitive poly $d(AC/TG)_{16}$ region due to faulty DNA mismatch repair results in in-frame variants which are detected as blue colonies (Fig. 12). The slippage events can be verified after plasmids rescued from *S. cerevisiae* (cf. preparation of yeast genomic DNA, Materials and Methods) are transformed into commercially available strains of *E. coli*. (Materials and Methods). The actual slippage event was identified by DNA sequencing using primer "PRS3 Forward" (cf. Materials and Methods). The procedure used was based on the chain-termination method (Sanger *et al.* 1977) using the ABI 373A sequencer (Applied Biosystems, Foster City, California, USA) (cf. Materials and Methods).

Sequencing of the repetitive region within the three assay plasmids revealed that the blue yeast colony on a plate corresponded to a slippage event at the DNA level. The most frequent slippage event was the loss of

one (AC) pair, the result of which shifted the coding region of the *PRS3'lacZ* fusion back into the correct reading frame giving rise to a functional gene product (Fig. 12). Once sequenced, resulting in-frame variants of the high copy, low copy integrative assay plasmids were measured for specific β-galactosidase activity in yeast (Fig. 12).

4(ii) Influence of certain food components on DNA instability and repair

4(iii) Screening for the potential negative influences of compounds on DNA stability and mismatch repair

As well as being able to screen for the protective (anticarcinogenic) effects of human dietary components on the stability of microsatellite DNA, it is also possible to screen for the negative (carcinogenic) influences of compounds using the preferred simple colour selection system in yeast according to the first aspect of the invention. When screening for carcinogens, a wild type yeast strain for mismatch repair is transformed with the high copy assay construct for increased sensitivity. This transformed strain is then grown in the presence or absence of carcinogens and the frequency of white to blue colonies arising from the different growth conditions are subsequently compared. Carcinogens increase the rate of white to blue transition as compared to "normal" growth conditions (cf. Fig. 1).

The intercalating agent ethidium bromide and its effect on microsatellite instability

Ethidium bromide - a known potent carcinogen causes frame-slippage through its action of intercalation. This and other intercalating agents preferentially target monotonic runs or alternating nucleotide sequences. As shown in Figure 17, at higher concentrations (6-7μg/ml) this agent increases the frequency of frame-slippage 300 fold compared to an untreated strain (cf. Figure 1).

Method

Yeast strain YN94-1 was transformed with pKa3-9(32 1) and plated onto selective medium (cf. Materials and Methods). A single colony was inoculated into 100ml SCD-Uracil (cf. Materials and Methods) and grown to mid-log phase. Culture was divided into 10 x 10ml sterile bottles. Cells were harvested, washed in 0.1 M potassium phosphate (KPP) buffer (pH 6.5) then resuspended in KPP buffer at a concentration of approximately 10^6 cells/ml. Ethidium bromide was then added to concentrations between 0 and 9μg/ml, and the cells were incubated at 30°C with agitation for approximately 7 hours. Following incubation the cells were washed with 0.1 M KPP buffer, diluted in 10ml H_2O and plated on SCD Uracil for single washed with 0.1 M KPP buffer, diluted in 10ml H_2O and plated on SCD Uracil for single colonies (4 plates for each ethidium bromide concentration). This experiment was repeated three times.

Polyamines and DNA stability

Polyamines - a group of flexible polycations are normal constituents of the cell and are essential for many cellular processes. They are found in high

concentrations in red meat, fish and vegetables. Under physiological conditions putrescine, spermidine and spermine are protonated and possess two, three and four positive charges respectively. Spermine, with its four positive charges binds two phosphate groups in each strand of the DNA helix. This spanning of the major and minor groove by spermine stabilises the DNA helix (Heby, O. & Persson, L., 1990).

The frequency of frame-slippage for the preferred centromere-based assay plasmid pKaCEN(32-1) of Example 1(i) has been determined in yeast strains disrupted in steps of polyamine synthesis (Fig. 11 and Table 2). Results so far indicate that polyamines are required for the stabilisation of DNA during replication and could therefore aid in the prevention of frame-slippage.

Table 2: Frequency of alteration in lengths of poly(AC) tracts in wild type yeast strains and polyamine mutants.

Strain	Relevant genotype	Tract location	Frequency of tract alteration	Frequency relative to wild type
2602	wild type	pKaCEN(32-1)	6.20×10^{-6}	1
Y359	*spe1*	pKaCEN(32-1)	1.02×10^{-4}	16.35
Y362	*spe1 spe2*	pKaCEN(32-1)	1.54×10^{-4}	24.79

A simple colour assay for monitoring spontaneous base-base mismatches in *S. cerevisiae*.

As described, the mismatch repair pathway of eukaryotes repairs both insertion/deletion mismatches and spontaneous base-base mismatches. Therefore, using the above yeast strains we can use our assay to monitor the genetic and dietary influences on spontaneous base/base mismatches in yeast.

All of the above mentioned yeast strains are derived from the wild type yeast YN94-1 available from Dr Michael Stark, Department of Biochemistry, University of Dundee, DD1 4HN, Dundee, Scotland, UK and Professor Michael Schweizer, Genetics & Microbiology Department, Institute of Food Research, Norwich Laboratory, Norwich Research Park, Norwich, NR4 7UA, England, UK and as such contain the *ade*2-1 point mutation which affects the biosynthesis of purine nucleotides (Figure. 18). The red pigment that accumulates in *ade*1 and *ade*2 mutants derives from an intermediate formed in reaction 5. This step involves the closure of an imidazole ring by phosphoribosylaminoimidazole synthetase to yield phosphoribosylaminoimidazole (AIR). Obviously, wild type *ADE2* cells do not accumulate this pigment and therefore remain white. Therefore, subsequent reversion of red *ade* - strains to wild type enables us to monitor the frequency of spontaneous mutations by a red to white colour transition. We have subsequently measured the basal frequency of red to white reversion in the various genetic backgrounds of this system (Table 3)

Table 3 Frequency of *ADE*$^+$ reversion in various yeast genetic backgrounds

Strain	Frequency of *ADE*$^+$ reversion	Rate relative to wild type
wild type	2.51×10^{-4}	1
msh2	1.60×10^{-2}	66
msh3	4.80×10^{-4}	2
msh6	3.57×10^{-3}	14
mlh1	1.87×10^{-2}	74.5
rth1	3.42×10^{-3}	13.6
pol30	6.30×10^{-3}	25

4(iii) **Industrial Applicability**

The identification of human dietary components that protect against DNA instability and therefore some types of cancer by use of the present invention will contribute to the scientific basis for a healthy diet. The simple blue/white colour test according to a preferred embodiment can be provided in kit-form or scaled up for use in the food or pharmaceutical industries.

Once a test substance is identified as containing "protective" factors, programmes can be undertaken to characterize and elucidate the mode of actin of the protective factor within the foodstuff. The results from the assay of the invention should be of enormous value to plant and crop breeders who wish to produce foodstuffs of greater nutritional value.

It has been observed that drug resistant human ovarian carcinoma cell lines aquire a mutator phenotype and a deficiency in *hMLH1* repair activity, with loss of expression of the *hMLH1* subunit occurring in 9/10 independently derived cisplatin resistant sublines (Hirst *et al.* 1997). This observation shows that loss of *hMLH1* expression is a frequent event in the development of drug resistance and support the involvement of mismatch repair in mediating the cytotoxic action of chemotherapeutically important drugs. The assays of the invention include yeast strains carrying this *mlh1* mutation and so could be used in an *in vivo* study looking at the effects of cytotoxic agents and subsequent resistance.

Table 1a Frequency of alteration in lengths of poly(AC)n tracts in wild type yeast strains and mismatch repair mutants

Strain	Relevant genotype	Tract	Reading-frame	Frequency of tract alteration	Frequency relative to wild type
<u>High copy 2 μ based assay constructs</u>					
YN94-1	wild type	$(AC)_{16}$	-1	1.35×10^{-4}	1
YN94-1	wild type	$(AC)_{41}(A)_2(C)_2(AC)_4A$	-1	1.16×10^{-2}	1
YN94-1	wild type	$(AC)_{11}$	+1	1.37×10^{-4}	1
YN94-1	wild type	$(AC)_{12}A$	+1	1.40×10^{-4}	1
YN94-1	wild type	$(AC)_{15}A$	+1	3.27×10^{-4}	1
YN94-1	wild type	$(AC)_{24}A$	+1	5.40×10^{-4}	1
YN94-1	wild type	$(AC_{1-3})_n$	-1	2.95×10^{-5}	1
YN97-10	*msh2*	$(AC)_{16}$	-1	6.80×10^{-2}	504
YN97-10	*msh2*	$(AC)_{12}A$	+1	7.41×10^{-2}	529
YN97-11	*mlh1*	$(AC)_{16}$	-1	1.14×10^{-1}	843
YN97-11	*mlh 1*	$(AC)_{12}A$	+1	8.50×10^{-2}	635
YN97-147	*msh3*	$(AC)_{16}$	-1	6.06×10^{-3}	45
YN97-147	*msh3*	$(AC)_{12}A$	+1	2.90×10^{-3}	21
YN97-150	*msh6*	$(AC)_{16}$	-1	1.20×10^{-3}	9
YN97-150	*msh6*	$(AC)_{12}A$	+1	1.40×10^{-3}	8
YN97-167	*rth1*	$(AC)_{16}$	-1	5.43×10^{-3}	40
YN97-167	*rth1*	$(AC)_{12}A$	+1	4.76×10^{-2}	340
YN98-3	*pol30-104*	$(AC)_{16}$	-1	9.47×10^{-3}	70
YN98-3	*pol30-104*	$(AC)_{12}A$	+1	3.03×10^{-3}	21

Table 1a, continued overleaf

Table 1a, continued

low copy centromere-based constructs

YN94-1	wild type	$(AC)_{16}$	-1	$1.40x10^{-5}$	1
YN94-1	wild type	$(AC)_{41}(A)_2(C)_2(AC)_4A$	-1	$1.72x10^{-4}$	1
YN94-1	wild type	$(AC)_{12}A$	+1	$1.65x10^{-5}$	1
YN94-1	wild type	$(AC)_{15}A$	+1	$2.27x10^{-5}$	1
YN94-1	wild type	$(AC)_{24}A$	+1	$4.40x10^{-5}$	1
YN97-10	*msh2*	$(AC)_{16}$	-1	$5.90x10^{-3}$	421
YN97-10	*msh2*	$(AC)_{41}(A)_2(C)_2(AC)_4A$	-1	-	-
YN97-10	*msh2*	$(AC)_{12}A$	+1	$4.29x10^{-3}$	260
YN97-10	*msh2*	$(AC)_{15}A$	+1	$5.25x10^{-3}$	231
YN97-10	*msh2*	$(AC)_{24}A$	+1	$9.19x10^{-3}$	209
YN97-11	*mlh1*	$(AC)_{16}$	-1	$6.14x10^{-3}$	439
YN97-11	*mlh1*	$(AC)_{12}A$	+1	$7.27x10^{-3}$	258
YN97-147	*msh3*	$(AC)_{16}$	-1	$1.15x10^{-3}$	82
YN97-147	*msh3*	$(AC)_{12}A$	+1	$6.11x10^{-4}$	22
YN97-150	*msh6*	$(AC)_{16}$	-1	$1.50x10^{-4}$	11
YN97-150	*msh6*	$(AC)_{12}A$	+1	$9.17x10^{-5}$	6
YN97-167	*rth1*	$(AC)_{16}$	-1	$2.72x10^{-4}$	20
YN97-167	*rth1*	$(AC)_{12}A$	+1	$4.30x10^{-3}$	260
YN98-3	*pol30-104*	$(AC)_{16}$	-1	$1.72x10^{-3}$	123
YN98-3	*pol30-104*	$(AC)_{12}A$	+1	$3.29x10^{-4}$	20

Table 1a, continued overleaf

Table 1a, continued

single copy integrated constructs

YN94-1	wild type	$(AC)_{16}$	-1	$<5.0x10^{-6}$	1
YN94-1	wild type	$(AC)_{12}A$	+1	$<5.0x10^{-6}$	1
YN97-10	*msh2*	$(AC)_{16}$	-1	$1.59x10^{-3}$	318
YN97-10	*msh2*	$(AC)_{12}A$	+1	$1.80x10^{-4}$	180
YN97-11	*mlh1*	$(AC)_{16}$	-1	$1.82x10^{-3}$	363
YN97-11	*mlh1*	$(AC)_{12}A$	+1	$2.19x10^{-3}$	438
YN97-147	*msh3*	$(AC)_{16}$	-1	$1.45x10^{-4}$	29
YN97-147	*msh3*	$(AC)_{12}A$	+1	$1.00x10^{-4}$	20

REFERENCES

Aaltonen, L.A., Peltomaki, P., Leach, F.S., Sistonen, P., Pylkkanen, L. (1993), *Clues to the pathogenesis of familiar colorectal cancer, Science*, **260**, 812-816.

Birnboim, H.C., and Doly, J. (1979) *A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research*, 7, 1513.

Carter, A.T., Narbad, A., Pearson, B.M., Beck, K.F., Logghe, M., Contreras, R., Schweizer, M. (1994). *Phosphoribosylpyrophosphate synthetase (PRS): a new gene family in Saccharomyces cerevisiae. Yeast,* **10**, 1031-1044.

Elble, R. (1992). *A simple and efficient procedure for transformation of yeast. Biotechniques*, **13**, 18-20.

Eshelman, J.R. and Markowitz, S.D. (1995). *Curr. Opin. Oncol.* **7**, 83-89.

Heby, O. and Persson, L. (1990). *Molecular genetics of polyamine synthesis in eukaryotic cells. TIBS.* **15**, 153-158.

Henderson, S.T., Petes, T.D. (1992). *Instability of simple sequence DNA in Saccharomyces cerevisiae. Mol. Cell. Biol.*, **12**, 2749-2757.

Mandel, M. and Higa, A. (1970). *Calcium dependent bacteriophage DNA infection. J. Mol. Biol.*, **53**, 154.

Maniatis, T., Fritsch, E.F., Sambrook, J. (1983). *Molecular cloning a laboratory manual, 7th Ed.* Cold Spring Harbour, N.Y.

Merlo, A., Mabry, E., Gabrielson, R., Vollmer, R., Baylin, S.B. (1994), *Frequent instability in primary small cell lung cancer. Cancer Res.* **54**, 2098-2100.

Miller, J.H. (1972). *Experiments in molecular genetics.* Cold Spring

Harbour Laboratory, Cold Spring Harbour, N.Y.

Mironov, N.M., Augelon, A.M., Potapova, G.I., Omori, Y., Gurbunov, O.V. (1994), *Alterations of (CA)n DNA repeats and tumour supressor genes in human gastric cancer. Cancer Res.* **54**, 41-43.

Modrich, P. and Lahue, R. (1996), *Mismatch repair in replication fidelity, genetic recombination and cancer biology. Annu. Rev. Biochem.*, **65**, 101-133.

Myers, A.M., Tzagoloff, A., Kinney, D.M., and Lusty, C. (1986). *Yeast shuttle and integrative vectors with multiple cloning sites suitable for construction of lacZ fusions. Gene,* **45**, 299-310.

Orth. K., Hung, A., Gezdar, A., Bowcock, J.M., Mathis. (1994), *Genetic instability in human ovarian cancer cell lines. Proc. Natl. Acad. Sci. USA*, **91**, 9495-9499.

Pearson, B.M. and McKee, R.A. (1992). *Rapid identification of Saccharomyces cerevisiae, Zygosaccharomyces bailii and Zygosaccharomyces rouxii. Int. J. Food Micro,* **16**, 63-67.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). *DNA sequencing with chain-termination inhibitors. Proc. Natl. Sci. USA*, **74**, 5463-5467.

Sikorski, R., and Hieter, P. (1989). *Genetics*, **122**, 17-19.

Strand, M., Prolla, T.A., Liskay, R.M., Petes, T.D. (1993). *Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature*, **365**, 274-278.

Strand, M., Earley, M.C., Crouse, G.F., Petes, T.D. (1995). *Mutations in the MSH3 gene preferentially lead to deletions within tracts of simple repetitive DNA in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA,* **92**, 10418-10421.

Umar, A. and Kunkel, T.A. (1996). *DNA-replication fidelity, mismatch*

repair and genome instability in cancer cells. Eur. J. Biochem., **238**, 297-307.

Weislander, L. (1979). *A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gell temperature agarose gels. Anal. Biochem,* **98**, 305.

Amin, S. N. and Holm, C. (1996) *In vivo* analysis reveals that the interdomain region of the yeast proliferating cell nuclear antigen is important for DNA replication and DNA repair. *Genetics*, 144, 479-493.

Alani, E. (1996). The *Saccharomyces cerevisiae* Msh2 and Msh6 proteins form a complex that specifically binds to duplex oligonucleotides containing mismatched DNA base pairs Mol. Cell. Biol. 16 p5604-5607.

Gietz, R. D. and Woods, R. A. (1994). High efficiency transformation with lithium acetate. In Johnson, J. R. (Ed.) Molecular Genetics of Yeast. A Practical Approach. IRL Press, Oxford, pp. 121-131.

Henderson, S. T., Petes, T. D. (1992). Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 12: 2749-2757.

Hirst, G. L. McIlwrath, A. J. Anthoney, D. A. Van Der Zee, A. and Brown, R. (1997). Loss of hMLH1 expression and drug resistance in ovarian tumour cells. Poster presentation, Ninth Tenovus - Scotland Symposium.

Johnson, R. E. Kovvali, G. K. Prakash, L & Prakash, S. (1995) Requirement of the yeast RTH1 5' to 3' exonuclease for the stability of simple repetitive DNA. Science 269: 238-239.

Johnson, R. E. Kovvali, G. K. Guzder, S. N. Amin, N. S. Holm, C. Habraken, Y. Sung, P. Prakash, L. Prakash, S. (1996). Evidence for involvement of yeast proliferating cell nuclear antigen in DNA mismatch repair. *J. Biol. Chem.* 271, 27987-27990.

Jones, E. W. and Fink, G. R. (1982) Amino acid and nucleotide biosynthesis. In Strathern, Jones and Broach (Ed.) The molecular biology of the yeast *Saccharomyces* - metabolism and gene expression). p.271-278. CHS Press.

Schweitzer, J. K. and Livingston, D. M. (1998). Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. Hum. Mol. Genet. 7,69-74

Shampay, J. S. Szostak, J. W. Blackburn, E. H. (1984) DNA sequences of telomeres maintained in yeast. Nature 310, 154-157.

Sia, E. A. Jinks-Robertson, S. & Petes, T. D. (1997). Genetic control of microsatellite stability. Mutation Research 383: 61-70.

Sia, E. A. Kokoska, R. J. Dominska, M. Greenwell, P. & Petes, T. D. (1997). Microsatellite instability in yeast: Dependence on repeat unit size and DNA mismatch repair genes. Mol. Cell. Biol. 17: 2851-2858

Sommers, C. H. Miller, E. J. Dujon, Prakash, S. & Prakash, L. (1995) Conditional lethality of null mutations in *RTH1* that encodes the yeast counterpart of a mammalian 5'- to 3'-exonuclease required for lagging strand DNA synthesis in reconstituted systems. J. Biol. Chem 270: 4193-4196.

Strand, M., Prolla, T. A., Liskay, R. M., Petes, T. D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365: 274-278.

Strand, M., Earley, M. C., Crouse, G. F., Petes, T. D. (1995). Mutations in the *MSH3* gene preferentially lead to deletions within tracts of simple repetitive DNA in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 92: 10418-10421.

Tishkoff, D. X. Filosi, N. Gaida, G. M. Kolodner, R. D. (1997). A novel mutation avoidance mechanism dependent on *S. cerevisiae RAD27* is distinct from DNA mismatch repair. Cell 88: 253-263.

Wach, A. Brachat, A. Pohlmann, R. and Philippsen (1994) New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. Yeast 10: 1793-1808.

Wierdl, M. Dominska and Petes, T. D. (1997) Microsatellite instability in yeast: Dependence on the length of the microsatellite. Genetics 146: 769-779.

Yanisch-Perron, C. Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene 33:103-119.

CLAIMS

1. An assay for testing the carcinogenic properties of a test substance comprising:

(i) introducing into cells a reporter gene expression vector comprising a repetitive DNA sequence which exhibits instability in cancer cells, whereby instability of the repetitive DNA sequence affects expression of the reporter gene;

(ii) exposing the resulting cells to the test substance; and

(iii) determining whether the test substance is carcinogenic or anti-carcinogenic by comparing the frequency of reporter gene expression in the resulting cells with the frequency of reporter gene expression in cells which have not been exposed to the test substance.

2. An assay as claimed in Claim 1 wherein the cells are eukaryotic cells.

3. An assay as claimed in Claim 1 or 2 wherein the cells are yeast or human cells.

4. An assay as claimed in Claim 1, 2 or 3 wherein the cells are defective in one or more repair processes, which defects lead to repetitive DNA sequence instability.

5. An assay as claimed in Claim 3 wherein the cells are human cells containing one or more mutations in the hMSH2 gene on chromosome 2 and/or the hMLH1 gene on chromosome 3.

6. An assay as claimed in Claim 3 wherein the cells are yeast cells which contain one or more mutations in the MLH1 gene and/or MSH2 gene and/or PMS1 gene.

7. An assay as claimed in any one of Claims 1 to 6 wherein the unstable repetitive DNA sequence comprises poly d(AC) and/or poly d(GT).

8. As assay as claimed in Claim 7 wherein the unstable repetitive DNA sequence comprises poly $d(AC)_n$ and/or poly $d(GT)_n$, wherein n is from 8 to 60, preferably from 16 to 32.

9. An assay as claimed in any one of Claims 1 to 8 wherein the repetitive DNA sequence exhibits instability in human colorectal cancer cells.

10. An assay as claimed in any one of Claims 1 to 9 wherein the reporter gene is selected from lacZ, URA3 or luciferase.

11. An assay as claimed in Claim 10 wherein the reporter gene comprises the lacZ gene.

12. An assay comprising:

(i) testing the carcinogenic properties of a test substance in yeast cells in an assay as claimed in Claim 3 or Claim 4 or any of their dependent claims; and

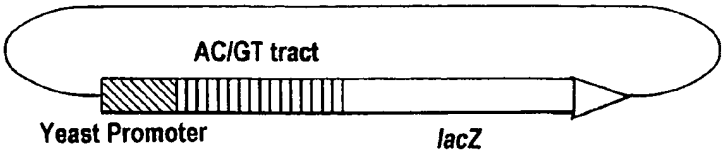
(ii) further testing the carcinogenic properties of the test substance in an assay in human cells, as claimed in Claim 3 and any of its dependent

claims.

13. An assay for testing the carcinogenic properties of a test substance comprising the steps of:

(i) screening a test substance using a reporter gene expression vector as claimed in any one of Claims 1 to 11; and

(ii) repeating the assay using cells containing a high, medium, low or single copy number reporter gene expression vector, the high, medium, low or single copy number vector being selected depending on the frequency of repetitive DNA instability measured in step (i).

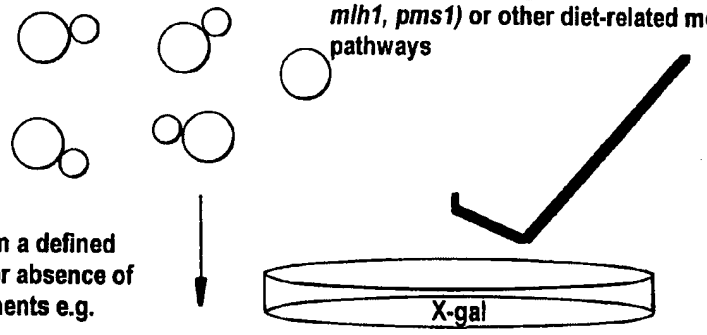


Assay plasmid for monitoring frame-slippage caused by faulty DNA mismatch repair in *Saccharomyces cerevisiae*

Fig. 1

Transformation into *Saccharomyces cerevisiae*

Yeast strains defective in the mismatch repair pathway (e.g. *msh2, msh3, msh6, mlh1, pms1)* or other diet-related metabolic pathways

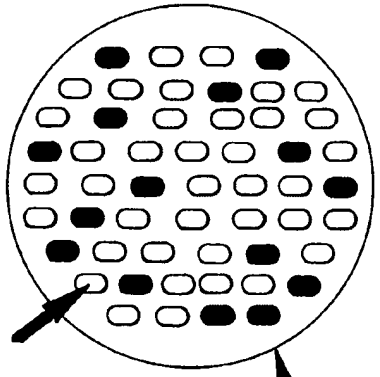


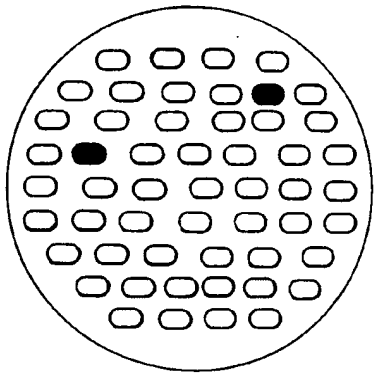
Yeast strains are grown on a defined medium in the presence or absence of protective dietary components e.g.
Broccoli extract
Polyamines: present in meat and vegetables
- known to play in role in DNA stability

Plates are incubated at 30°C to allow for growth and colour development

- Protective dietary components

+Protective dietary components





White yeast colony containing original out-of-frame yeast/*lacZ* fusion assay plasmid

Frame-slippage within the poly d(GT/CA) tract gives rise to a functional in-frame variant of the yeast/*lacZ* gene fusion. Faulty mismatch repair results in a white to blue transition

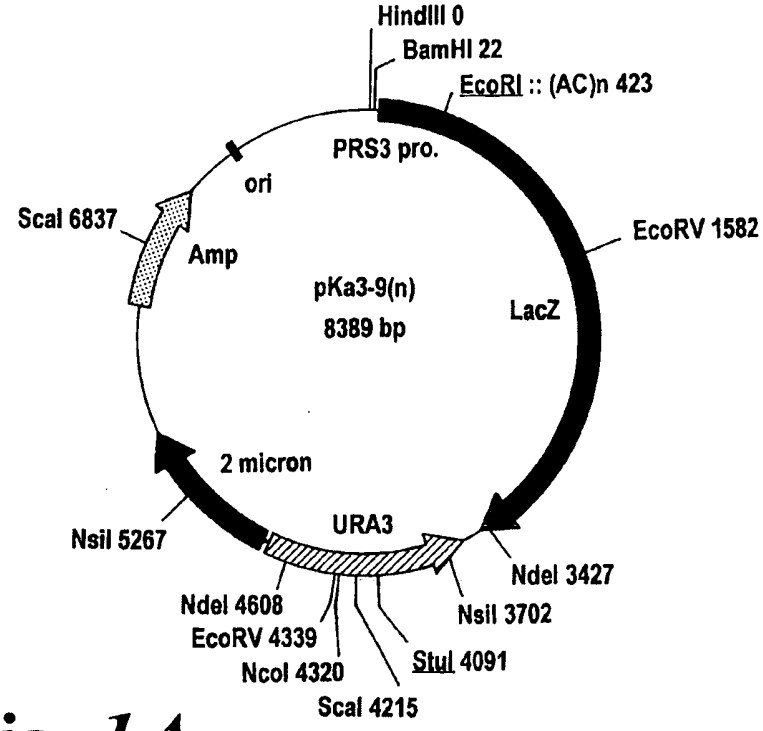

HindIII 0
BamHI 22
EcoRI :: (AC)n 423
PRS3 pro.
ori
ScaI 6837
Amp
EcoRV 1582
pKa3-9(n)
8389 bp
LacZ
2 micron
NsiI 5267
URA3
NdeI 3427
NdeI 4608
NsiI 3702
EcoRV 4339
StuI 4091
NcoI 4320
ScaI 4215

Fig. 1A

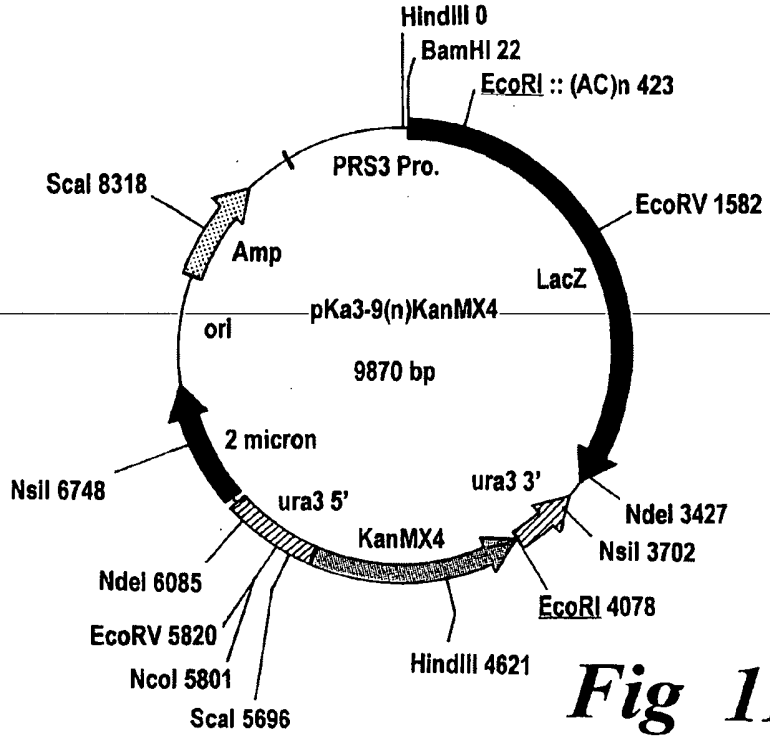
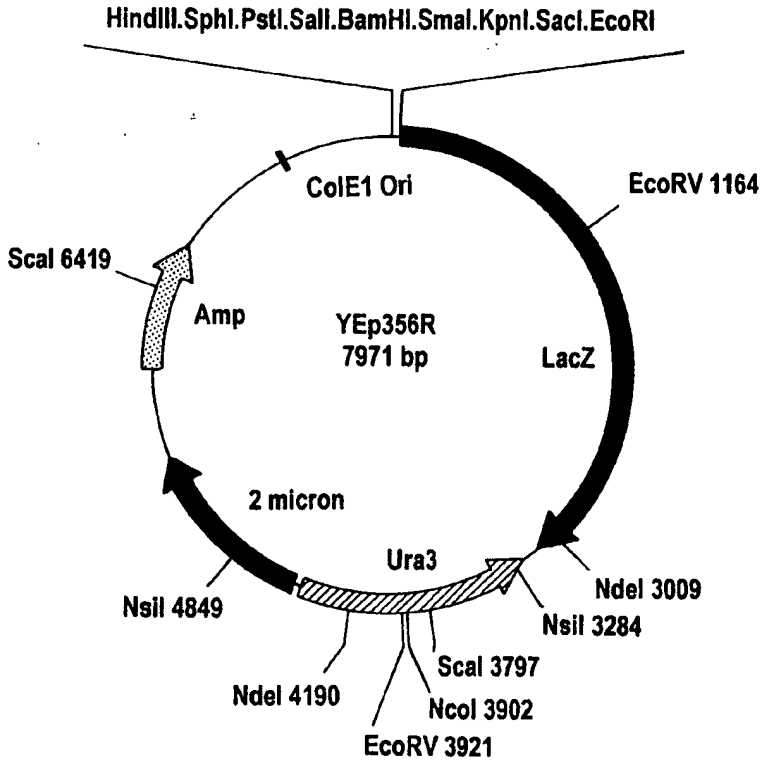

HindIII 0
BamHI 22
EcoRI :: (AC)n 423
PRS3 Pro.
ScaI 8318
EcoRV 1582
Amp
LacZ
pKa3-9(n)KanMX4
ori
9870 bp
2 micron
NsiI 6748
ura3 3'
ura3 5'
NdeI 3427
KanMX4
NsiI 3702
NdeI 6085
EcoRI 4078
EcoRV 5820
HindIII 4621
NcoI 5801
ScaI 5696

Fig 1B



YEp356R (Myers et al. 1986). A yeast/ E. coli shuttle vector suitable for fusing yeast promoter and coding sequences to the *lacZ* gene of E. coli. The 2 micron sequences ensure a high copy number of 25-100 per cell. The Amp^R and *URA3* genes allow selection in E. coli and yeast respectively.

Fig 2

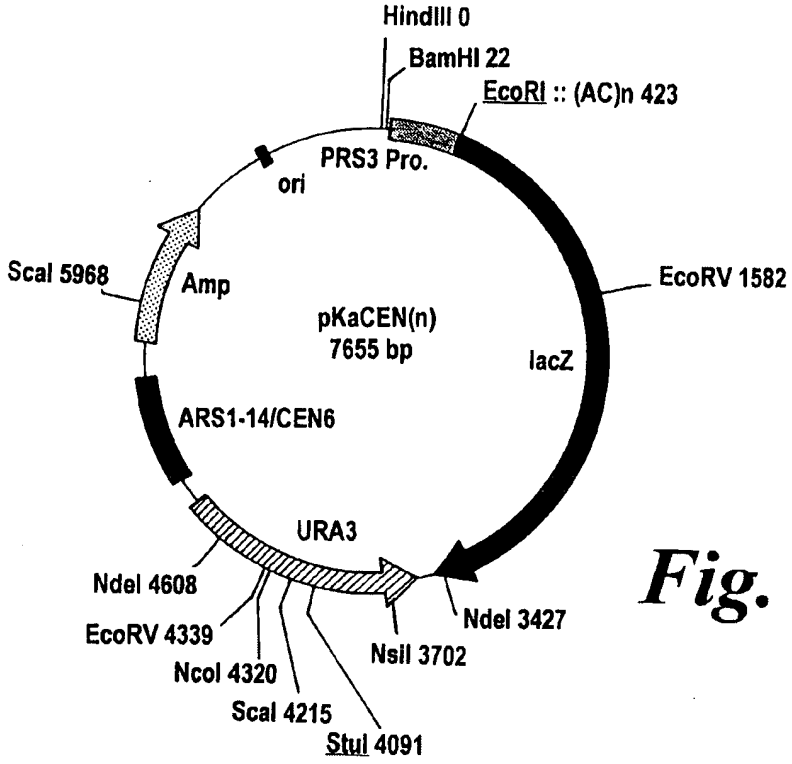

HindIII 0
BamHI 22
EcoRI :: (AC)n 423
PRS3 Pro.
ori
ScaI 5968
Amp
EcoRV 1582
pKaCEN(n)
7655 bp
lacZ
ARS1-14/CEN6
URA3
NdeI 4608
EcoRV 4339
NcoI 4320
ScaI 4215
StuI 4091
NdeI 3427
NsiI 3702

Fig. 2A

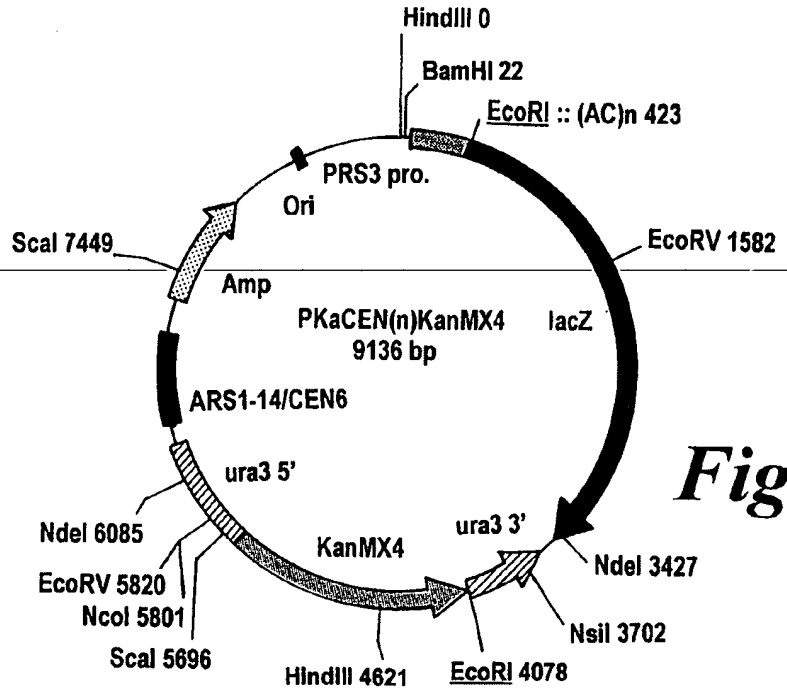

HindIII 0
BamHI 22
EcoRI :: (AC)n 423
PRS3 pro.
Ori
ScaI 7449
EcoRV 1582
Amp
PKaCEN(n)KanMX4
9136 bp
lacZ
ARS1-14/CEN6
ura3 5'
NdeI 6085
ura3 3'
KanMX4
NdeI 3427
EcoRV 5820
NcoI 5801
NsiI 3702
ScaI 5696
HindIII 4621
EcoRI 4078

Fig. 2B

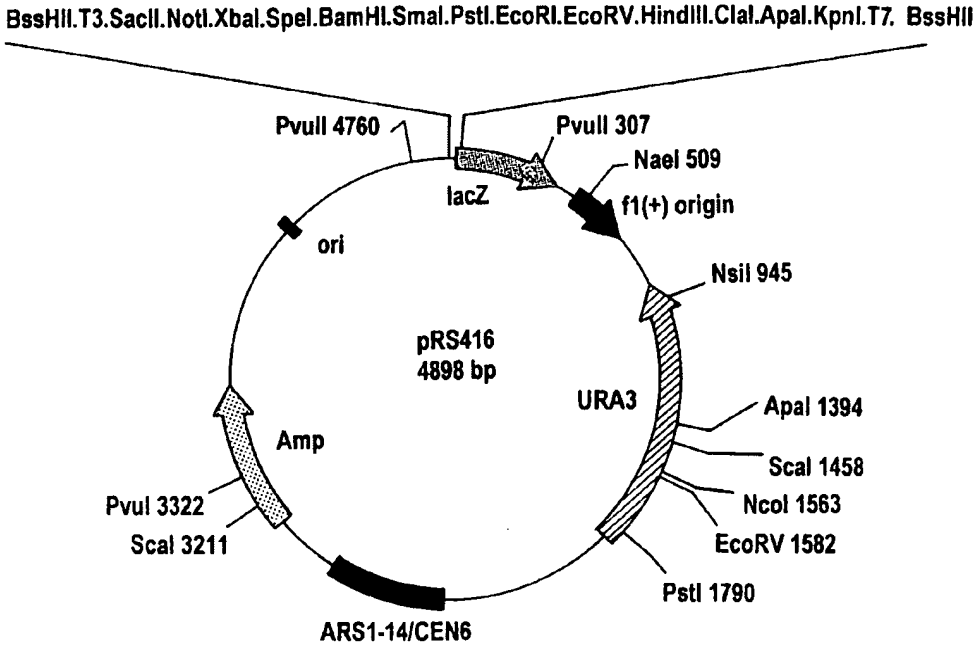


Fig. 3 pRS416 (Sikorski, R. & Hieter, P. (1989). A low copy number vector containing sequences from yeast centromere VI.

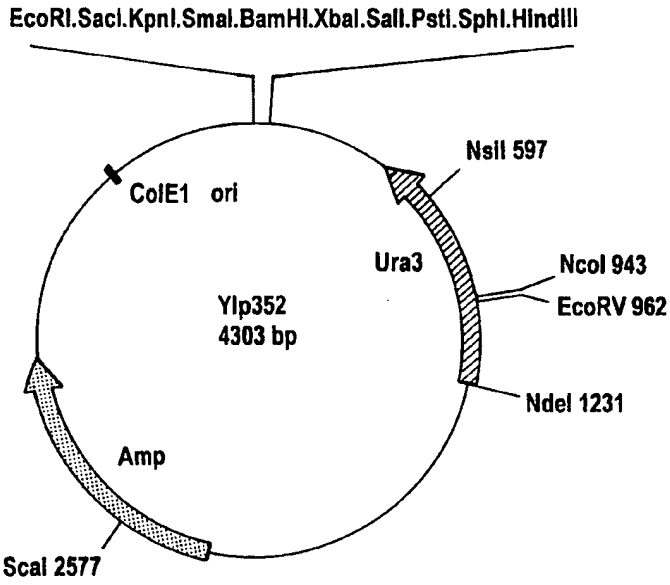
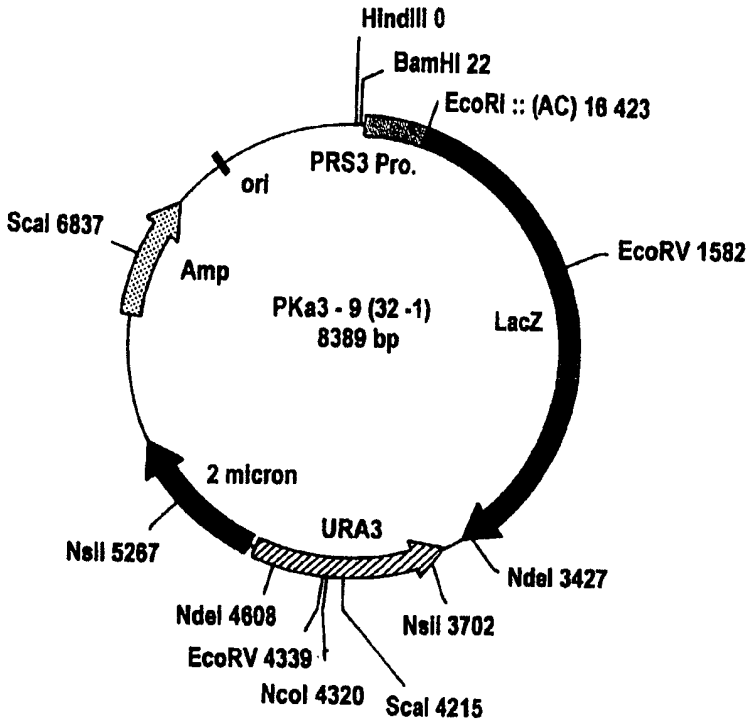


Fig. 4 YIp352 (Myers et al. 1986) A vector lacking a yeast origin of replication, therefore must be integrated into the yeast genome for stable maintenance.

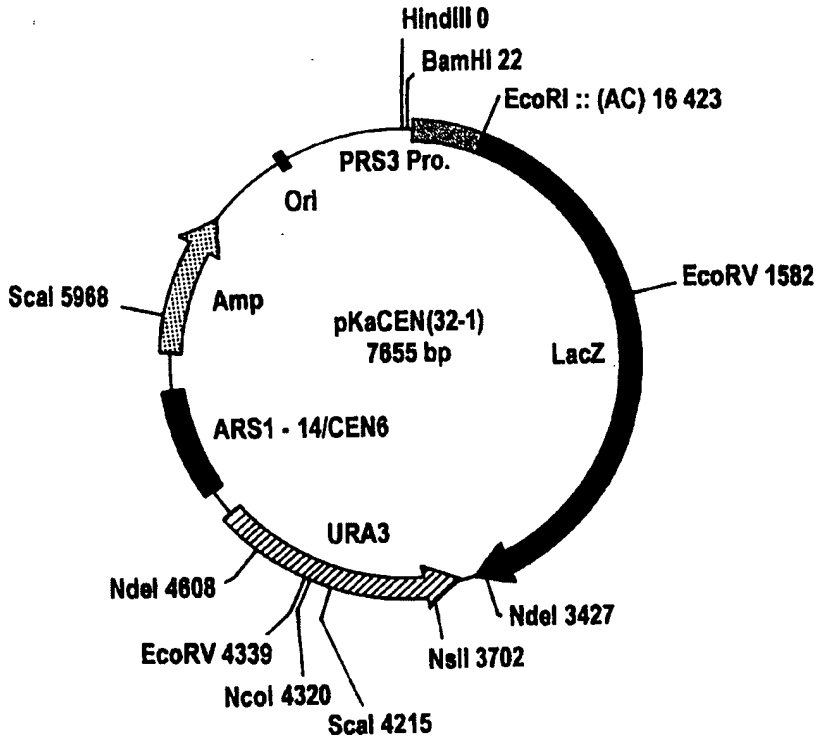
Fig. 5 pKa3 - 9 (32-1) (High copy construct, see text for details)



pKaCEN(32-1) (low copy construct)

Construction of this low copy centromere-based plasmid involved removal of the *PRS3'/lacZ* fusion containing the poly d(AC/TG)$_{16}$ tract from pKa3-9(32-1) on a 3680 bp *Bam*HI/*Nsi*I fragment (cf. Fig. 5) Initially pKa3-9(32-1) was digested to completion with *Bam*HI and *Nsi*I in restriction buffer REact™3 (Materials & Methods). KCl was added to adjust the salt concentration to that of REact™6 and digestion with *Sca*I was then carried out to eliminate incorporation of the smaller *Bam*HI/*Nsi*I fragment (Fig.). The resulting 3680 bp fragment containing the *PRS3 -'lacZ* fusion was gel purified (cf. Materials & Methods). pRS416 was also digested to completion with *Nsi*I and *BamH*I in buffer REact™3 with the aim of removing th existing *lacZ* region on a 923 bp fragment. The resulting 3975 bp vector band containing the ARS and

Fig. 6 **pKaCEN(32-1) (low copy construct, see text for details)**



pKaINT(32-1) (single copy integrative construct)

Construction of pKaINT(32-1) involved removal of the *PRS3'/lacZ* fusion containing the poly $d(AC)_{16}$ tract from pka3-9(32-1) on a *Bam*HI/*Nsi*I fragment. This fragment was subsequently cloned into *Bam*HI/*Nsi*I digested integrative plasmid YIp352 (cf. Fig. 4, Materials & Methods (Myers *et al.* 1986)). Before pKaINT(32-1) could be integrated into the yeast genome it was linearised by restriction at its unique

pKalNT(32-1) (single copy integrative construct, see text for details)

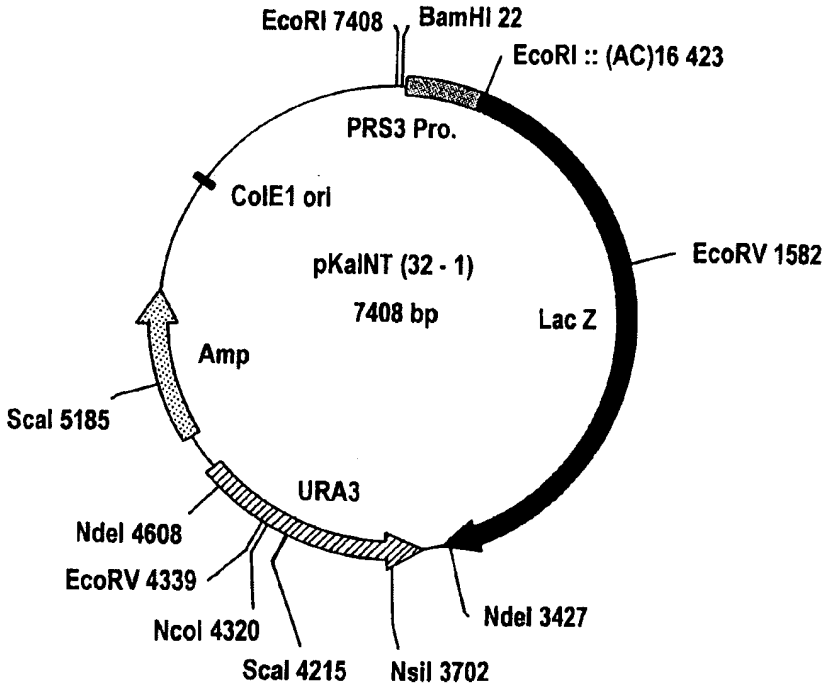
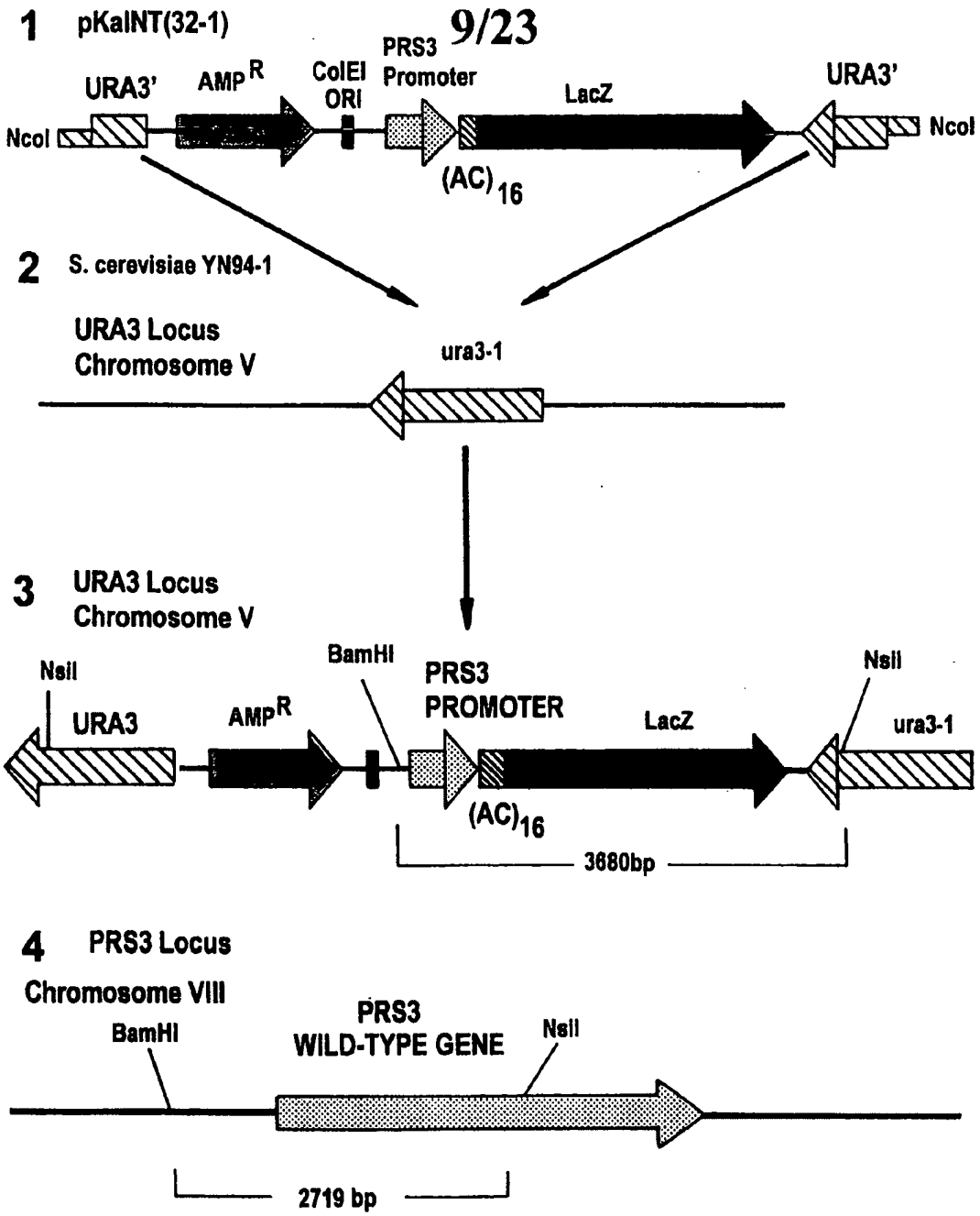
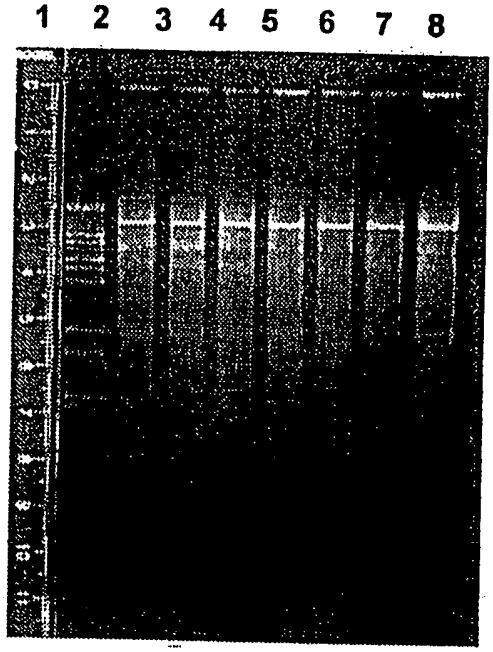


Fig. 7



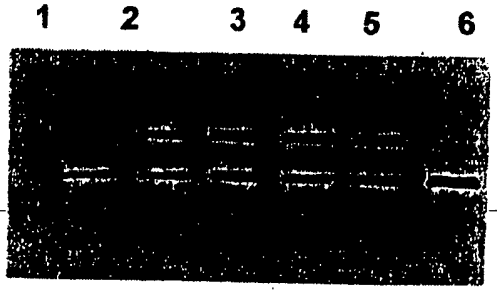
Scheme illustrating the integration of pKaINT(32-1) at the *URA3* locus of *S.cerevisiae* YN94-1. pKaINT(32-1) is linearised with *NcoI* (1) and integrated at the *URA3* locus on chromosome V through the process of homologous recombination (2). One wild type *URA3* and one mutant *ura3-1* are therefore generated(3). Finally, two restriction fragments will be detected when *BamHI/NsiI* digested yeast DNA is probed with the *PRS3* promoter region; firstly a fragment of 3680 bp containing the integrated *lacZ* fusion and secondly a 2719 bp fragment from the wild-type *PRS3*. (3 & 4).

Fig. 8



1% agarose gel showing *Bam*HI/*Nsi*I digests of genomic DNA of *S. cerevisiae* YN94-1.Lane (1) λ *Bst*EII marker, the sizes of the fragments are: 14140, 7242, 6369, 5687, 4822, 4324, 3675, 2323, 1929, 1371, 1264, 702, 224 & 117 bp respectively. Lane (2) *Bam*HI/*Nsi*I digested YN94-1 DNA control. Lanes (3-7) *Bam*HI/*Nsi*I digested DNA of *URA+* transformants.

Fig. 9



Corresponding Southern blot of the digests of genomic DNA shown in figure 9 . Lane (1)*Bam*HI/*Nsi*I digested control YN94-1 DNA. Only one 2719 bp band is seen corresponding to the wild type *PRS3* gene on yeast chromosome VIII. Lanes (2-7) *Bam*HI/*Nsi*I digests of genomic DNA containing integrated pKaINT(32-1). The gel was probed with a 423 bp *Eco*RI fragment derived the*PRS3* promoter. Two bands are clearly seen; one 2719 bp fragment corresponding to the wild type *PRS3* gene on chromosome VII and a larger 3680 bp fragment corresponding to the integrated vector at the *URA3* locus on chromosome V.

Fig 10

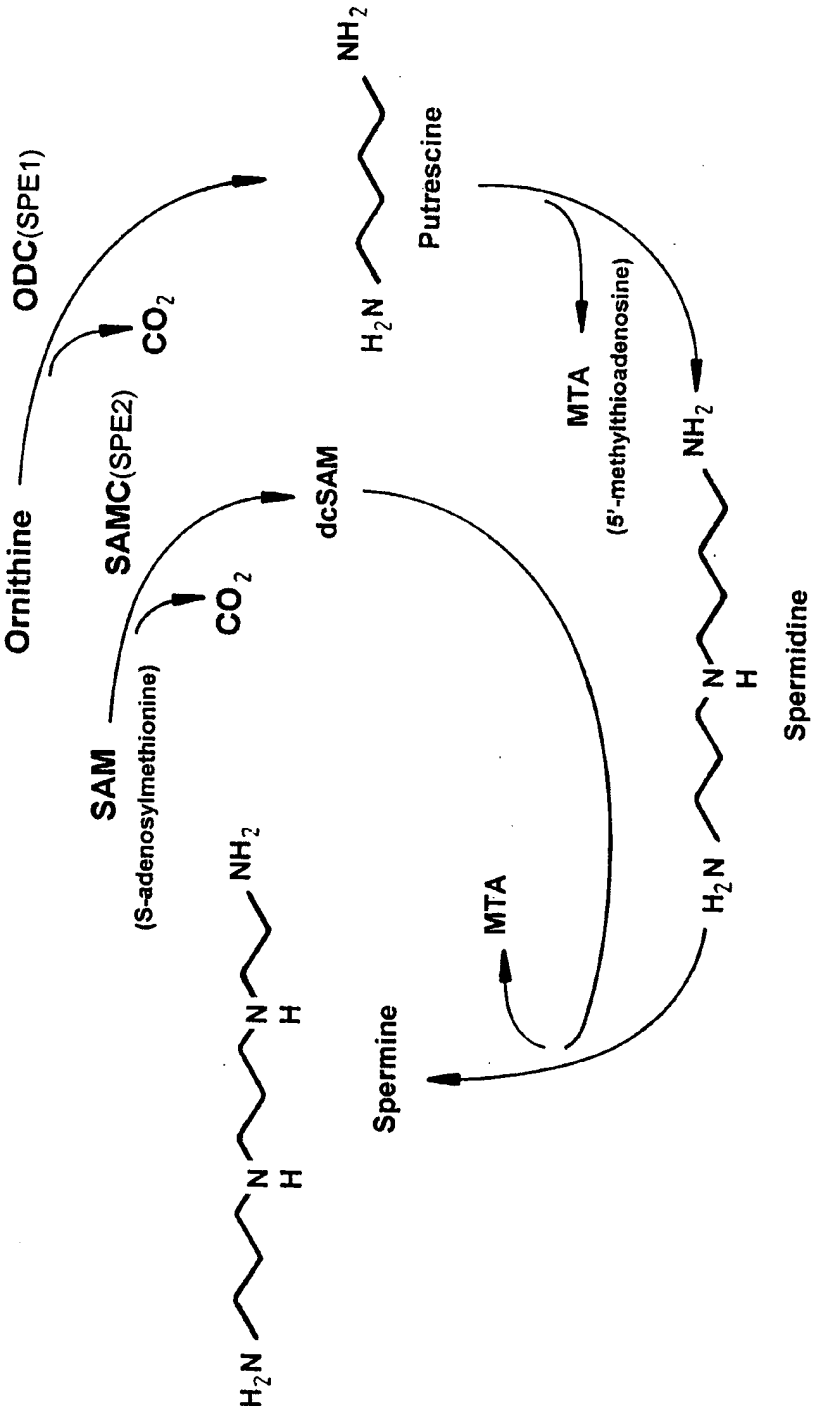


Fig. 11 Synthetic pathway for the polyamines putrescine, spermidine and spermine in eukaryotes

Schematic representation of frame-slippage in the three assay constructs. Scheme shows *PRS3'/lacZ* construct with repetitive tract in (AC) orientation with respect to the coding strand. Specific activity for β-galactosidase is shown for the out-of-frame construct and the in-frame variant where frame slippage has occured.

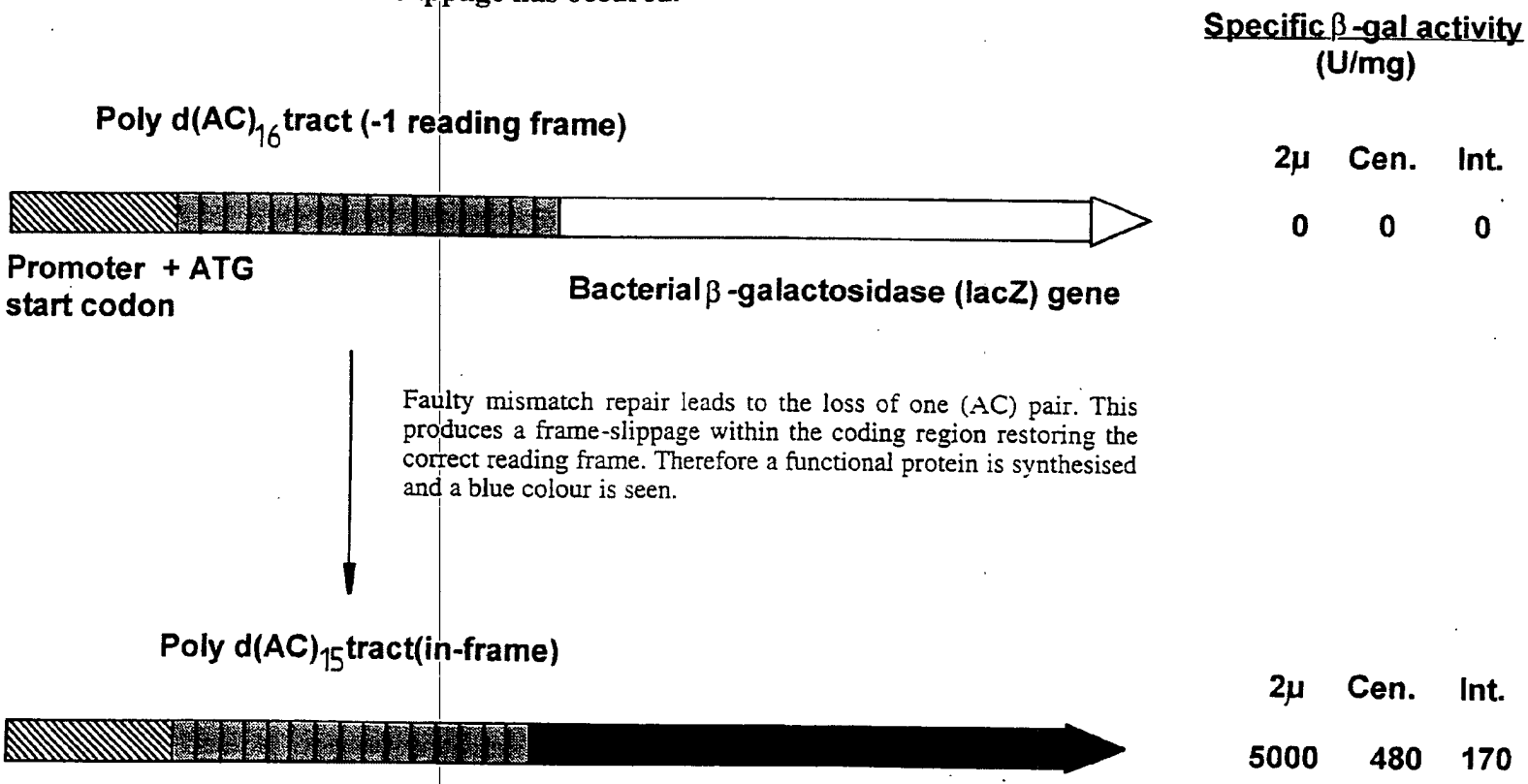


Fig. 12

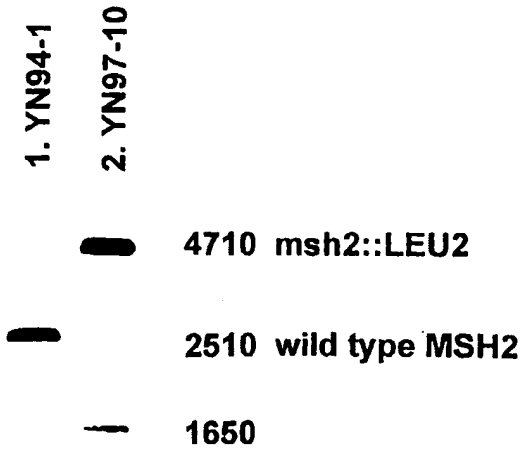


Fig. 13A

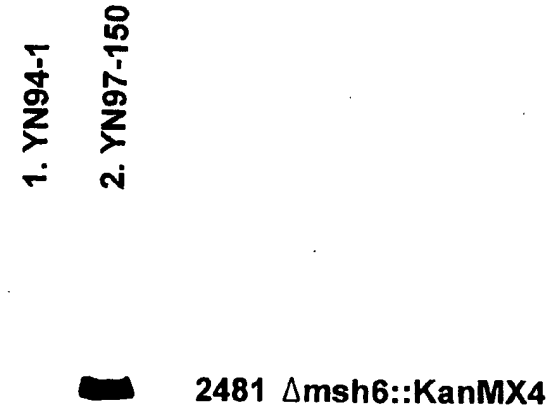
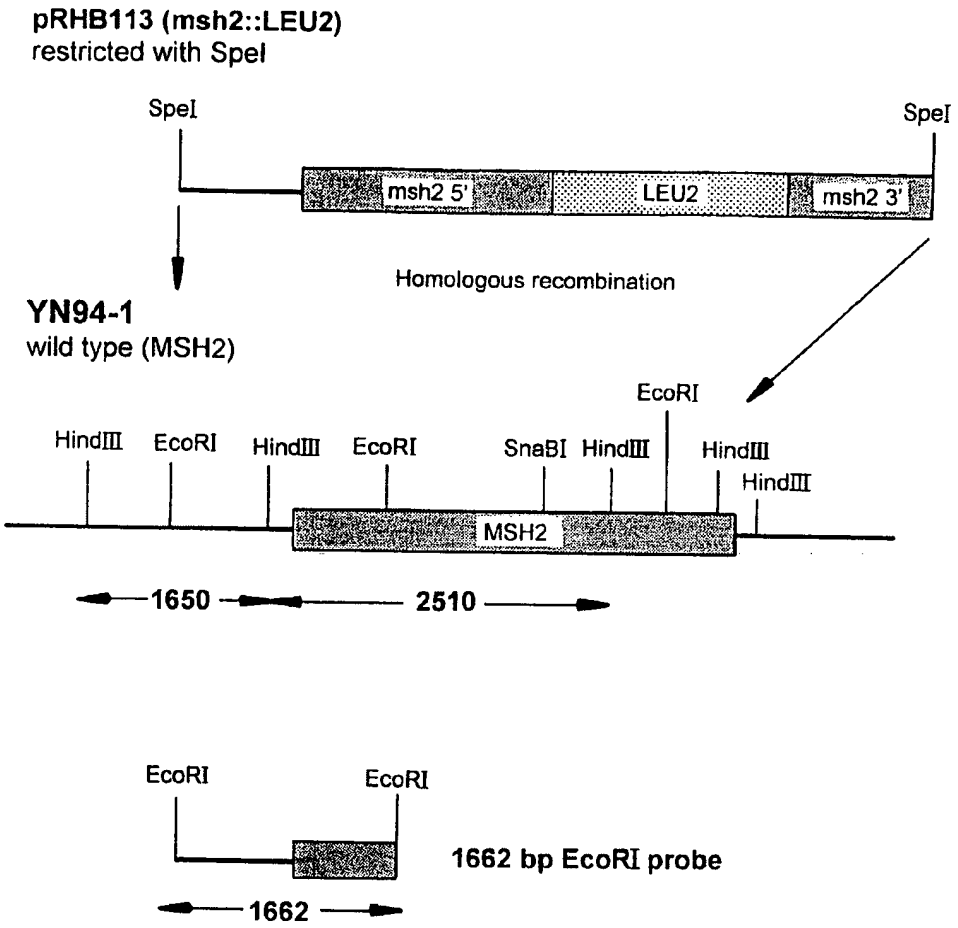


Fig. 15A



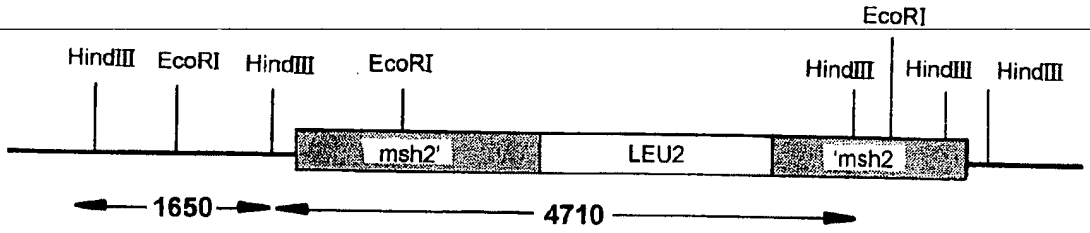


Fig 13B

South rn blot analysis of YN97-10 (msh2::LEU2)

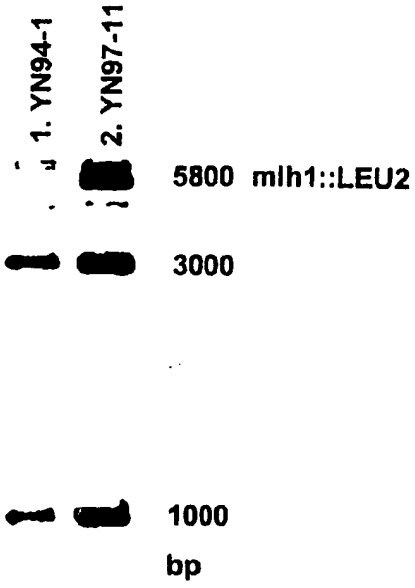


Fig. 14

Southern blot analysis of YN97-11 (mlh1::LEU2)

Lane (1) *Eco*RV digested control YN94-1 DNA. Two bands are seen (approximately 1000 and 3000 bp) corresponding to the wild type *LEU2* gene on chromosome III. Lane (2) *Eco*RV digests of yeast genomic DNA containing disrupted *MLH1*. The gel was probed with a 2200 bp *LEU2* fragment. Three bands are seen in the disruptant, two corresponding to the wild type *LEU2* on chromosome III and the other larger band corresponding to the *LEU2* gene used to disrupt *MLH1* on chromosome XIII (approximately 5800 bp).

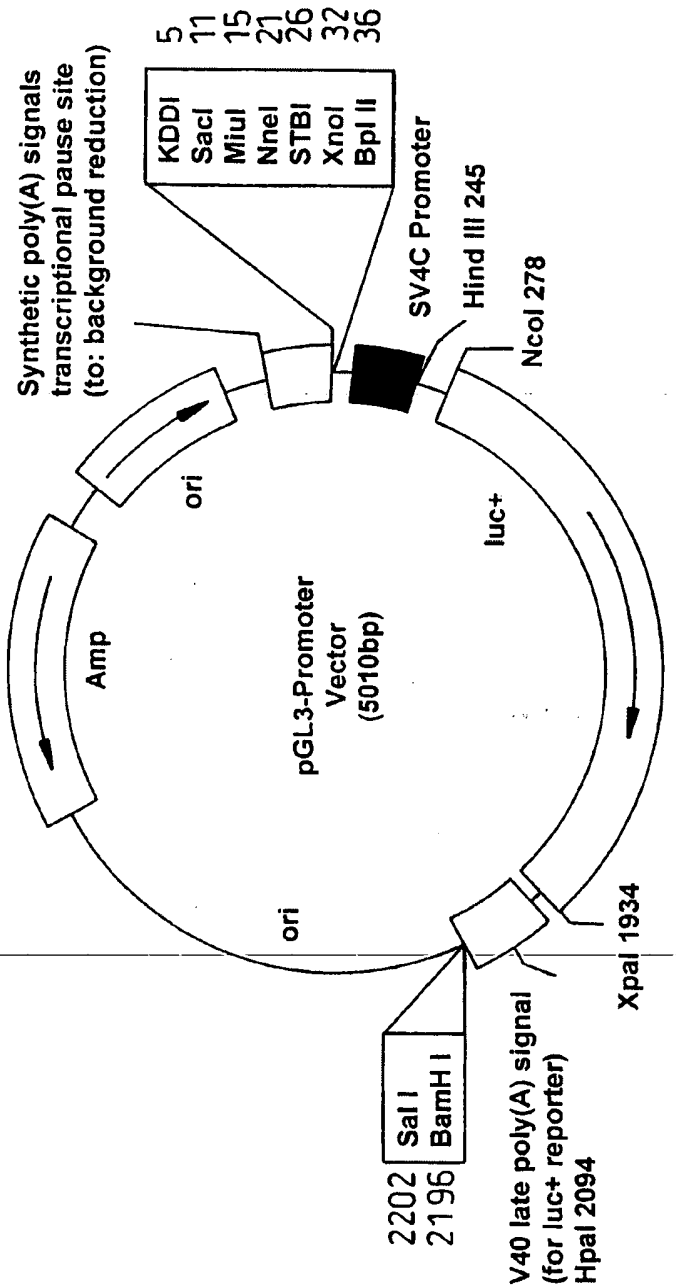


Fig. 15

pGL3-Promoter Vector (Promega-Biotec)
This vector contains an SV40 promoter upstream of the luciferase gene.

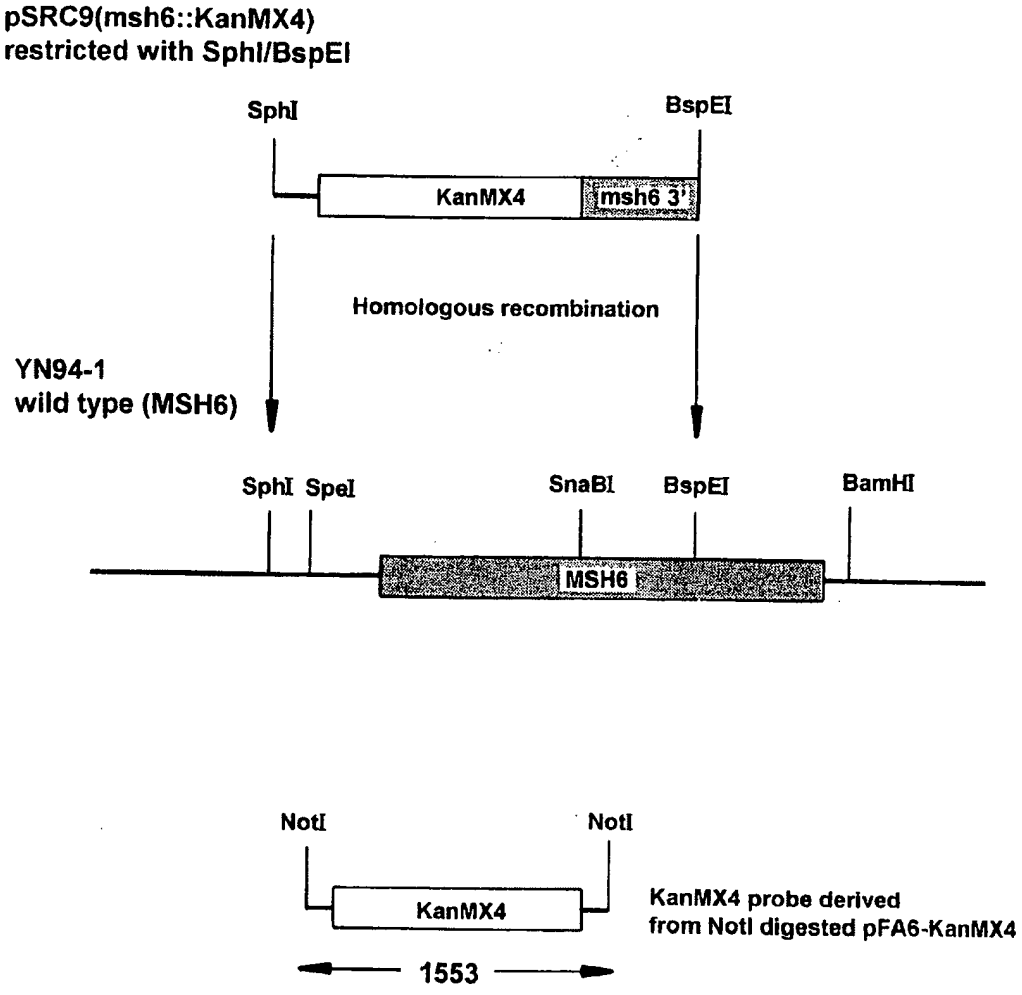


Fig 15B

South rn blot analysis of YN97-150 (msh6::KanMX4)

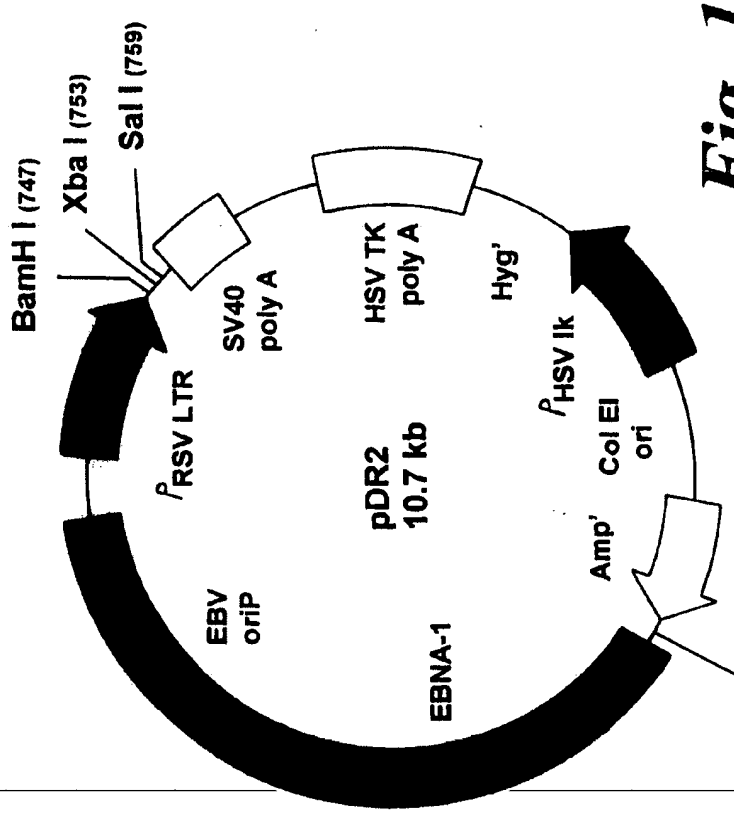


Fig. 16

pDR2

(Clontech Laboratories,Inc. 1020 East Meadow Circle, Palo Alto, California 94303-4230 USA)

A unique Epstein-Barr virus based vector for stable expression of DNA in human host cells. The EBV origin (oriP) confers stable episomal maintenance (10-20 copies per cell) to the vector when activated by the EBV nuclear antigen-1 (EBNA-1).

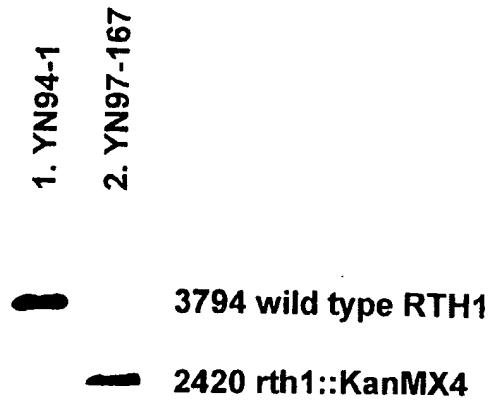


Fig. 16A

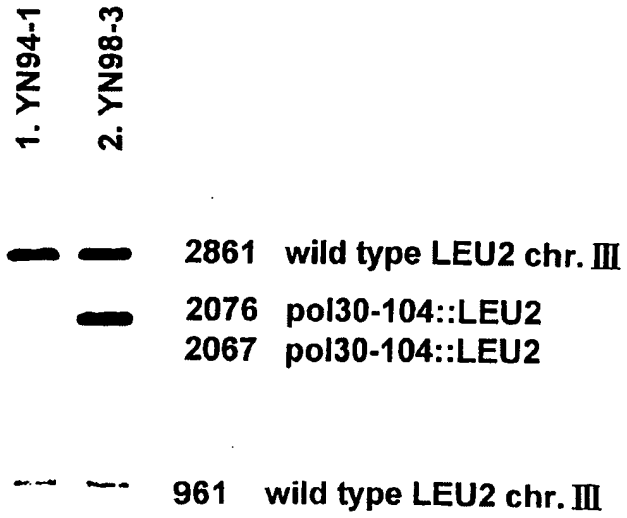


Fig 17A

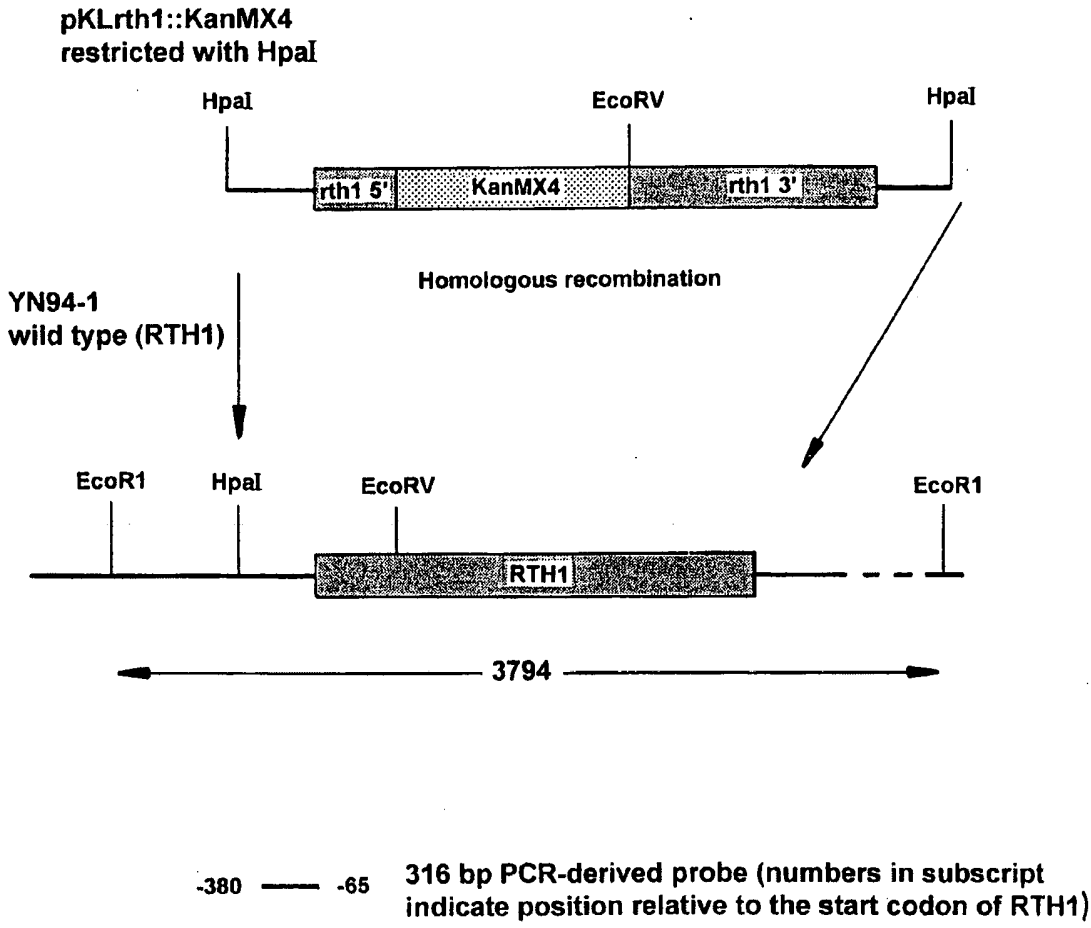


Fig 16B

Southern blot analysis of YN97-167 (rth1::KanMX4)

The effect of ethidium bromide on the stability of pKa3-9(32-1)

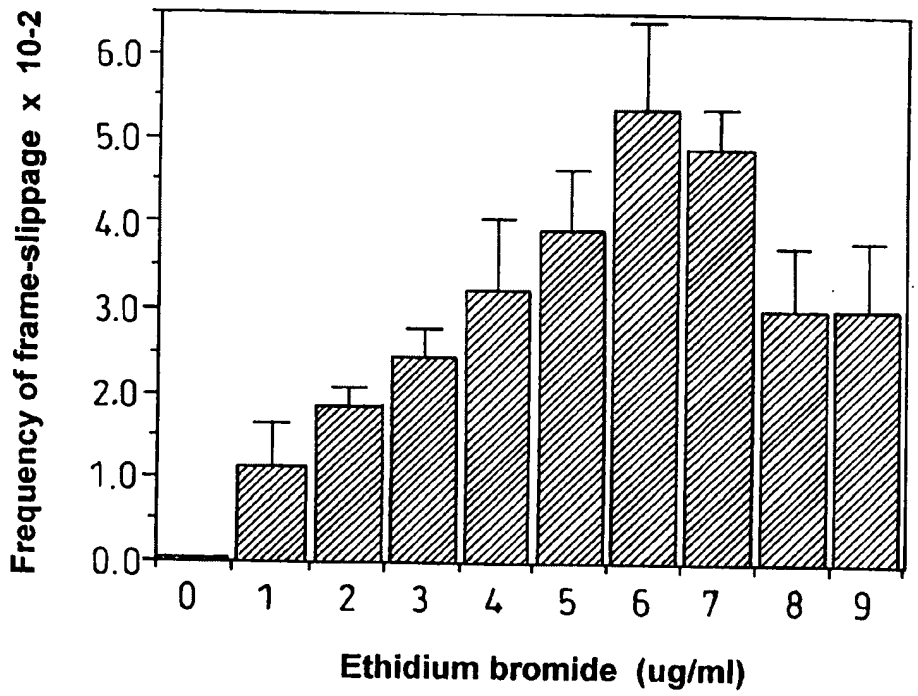


Fig. 17

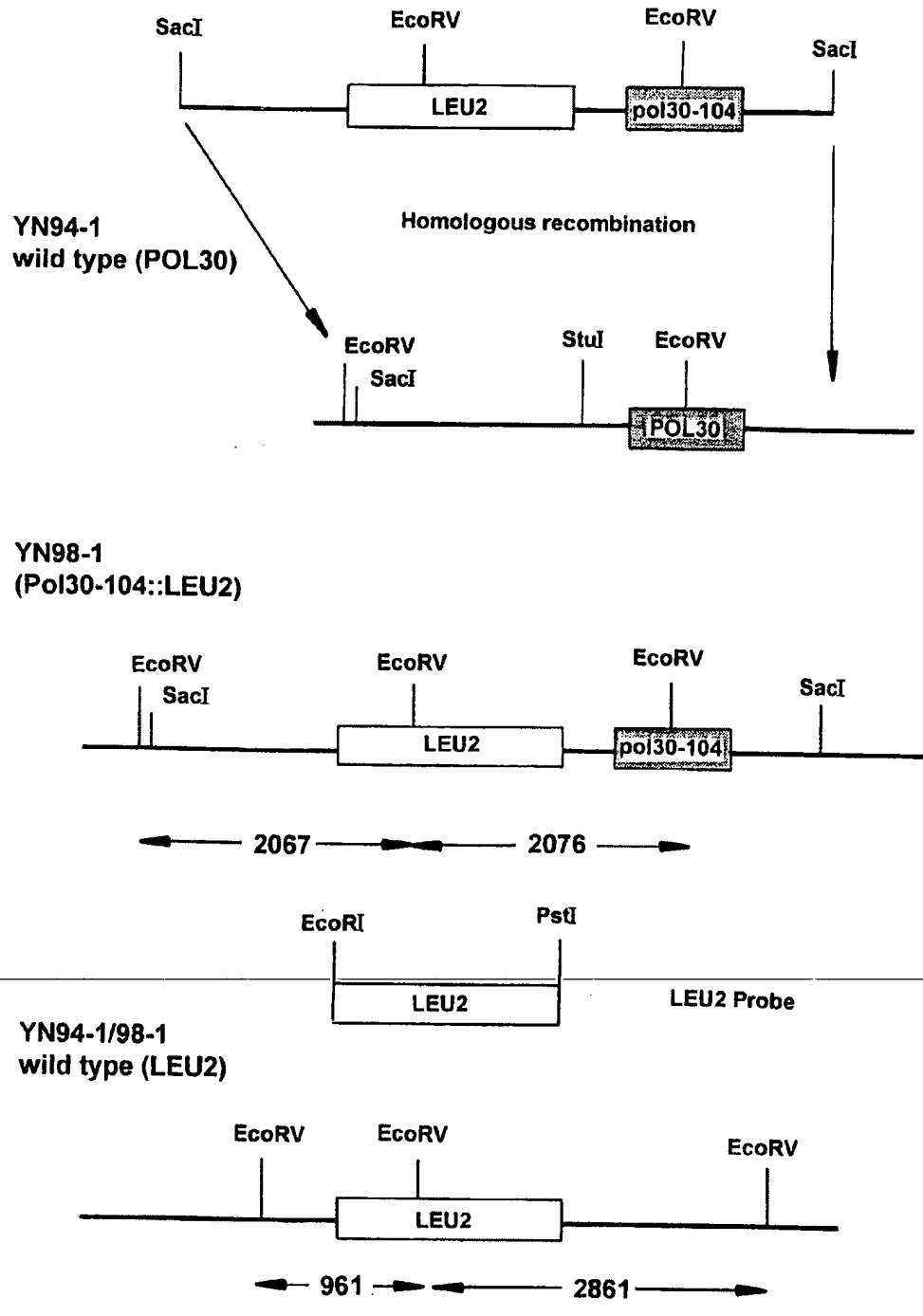


Fig 17B

South rn blot analysis of YN98-3 (pol30-104::LEU2)

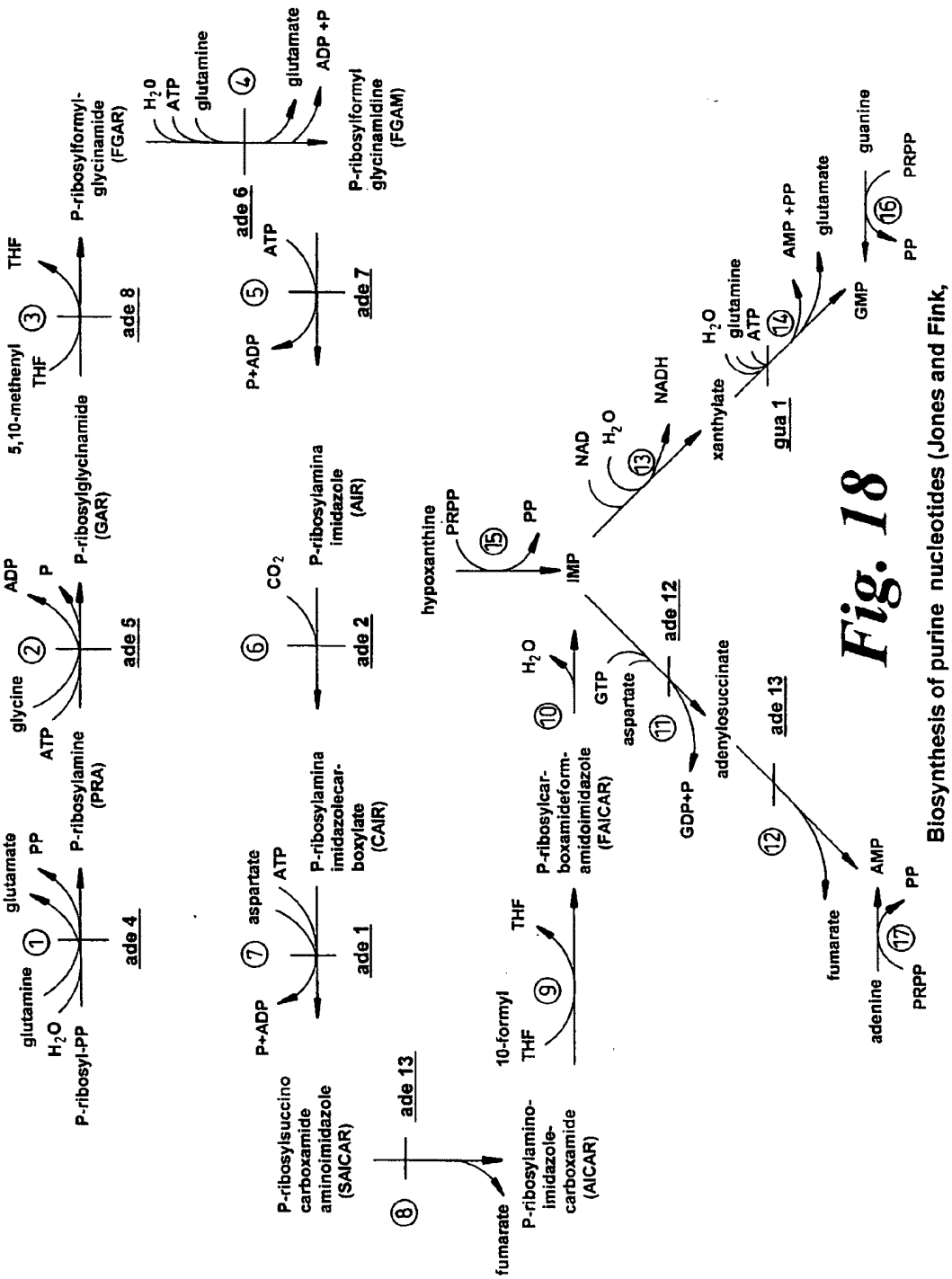


Fig. 18

Biosynthesis of purine nucleotides (Jones and Fink, 1982)

INTERNATIONAL SEARCH REPORT

Inter ›nal Application No
PCT/GB 98/00869

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68 G01N33/50

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimumdocumentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	STRAND ET AL.: "DESTABILIZATION OF TRACTS OF SIMPLE REPETITIVE DNA IN YEAST BY MUTATIONS AFFECTING DNA MISMATCH REPAIR" NATURE, vol. 365, 16 September 1993, pages 274-276, XP002004657 cited in the application see the whole document ---	1-13
Y	HENDERSON AND PETES: "INSTABILITY OF SIMPLE SEQUENCE DNA IN SACCHAROMYCES CEREVISAE" MOL. AND CELL. BIOLOGY, vol. 12, no. 6, 1992, pages 2749-2757, XP002075025 cited in the application see the whole document ---	1-13
	-/--	

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publicationdate of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of theinternational search	Date of mailing of the international search report
20 August 1998	03/09/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hagenmaier, S

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/GB 98/00869

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 0 258 899 A (MEDICAL RES COUNCIL) 9 March 1988 see the whole document	1-13
A	WO 89 05864 A (UNIV PRINCETON) 29 June 1989 see the whole document	1-13

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte onal Application No
PCT/GB 98/00869

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0258899 A	09-03-1988	AU 7793587 A	10-03-1988
		JP 63309194 A	16-12-1988
WO 8905864 A	29-06-1989	AU 2900389 A	19-07-1989
		EP 0390857 A	10-10-1990
		JP 3502882 T	04-07-1991